Bayesian learning of regularized Gaussian graphical networks



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

By

Jarod Mark Smith

14016665

Supervisor: Prof. Mohammad Arashi

Co-supervisor: Prof. Andriëtte Bekker

Submitted in fulfilment of the requirements for the degree
Philosophiae Doctor (Mathematical Statistics)

in the Faculty of Natural and Agricultural Sciences
Department of Statistics
University of Pretoria

January 2024

© University of Pretoria

© University of Pretoria

To Micaela, Ninni, Gary, Edelweiss, Tania, Victor, Jason, Monica, Ouma Liz, Ouma Edel, Oupa Smithy and Oupa Louis with love.

The measure of an education is that you acquire some idea of the extent of your ignorance.

Christopher Hitchens

I'd take the awe of understanding over the awe of ignorance any day.

Douglas Adams

© University of Pretoria

© University of Pretoria

Declaration

I declare that the dissertation, which I hereby submit for the degree *PhD Mathematical Statistics* at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

By
Jarod Mark Smith
14016665
January 2024

© University of Pretoria

© University of Pretoria

Acknowledgements

Without the various contributors listed below, this work would not have been able to stand as it is today.

- My home, Micaela: Your unwavering support, boundless patience, encouragement, and above all, your love, have been the cornerstones of my journey through this pursuit. You have not only kept my feet firmly planted on the ground but also lifted my spirits during the most challenging times.
- Prof. Mohammad Arashi and Prof. Andriëtte Bekker: Thank you for your unwavering support throughout the research process. Your guidance and expertise have been instrumental in shaping, not only this dissertation but also my personal and academic growth.
- Dr. Mattheüs Theodor Loots: I am deeply grateful for the gift of your friendship, along with your tireless listening and continuous advice in our discussions. You have not only helped me think critically and lifted me in times of self-doubt but have also been a source of solace during the heartaches of life. Thank you from the bottom of my heart.
- Jean-Pierre Stander: Our PhD journeys have been an epic adventure filled with highs and lows. Thanks for always sharing a laugh when we needed it most.
- Wilma van der Westhuizen: Thank you for your review, your attention to detail greatly improved the readability of this dissertation.
- Prof. Hao Wang: Thank you for the Bayesian graphical lasso, which served as the catalyst and bedrock for my dissertation. Upon revisiting your research, my appreciation for its elegance grew with each reading.
- Examiners: Thank you for reviewing this work, your attention to detail greatly improved the presentation, scientific justification and readability of this dissertation.
- This work was based upon research supported in part by the National Research Foundation (NRF) of South Africa, Ref.: SRUG190308422768 grant No. 120839. The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF. Mohammad Arashi's work is based on the research supported in part by the Iran National Science Foundation (INSF) grant No. 4015320.

© University of Pretoria

© University of Pretoria

Abstract

The advancement of digitisation in various scientific disciplines has generated data with numerous variables. Gaussian graphical models (GGMs) offer a convenient framework for analysing and interpreting the conditional relationships among these variables, with network inference relying on estimating the precision matrix within a multivariate Gaussian framework. Two novel Bayesian shrinkage methods are proposed for the estimation of the precision matrix. The first develops a Bayesian treatment of the frequentist alternative ridge precision estimator with the common ℓ_2 penalty, allowing for networks that are not necessarily highly sparse. The second caters for diverse sparsity by enabling both ℓ_1 and ℓ_2 based shrinkage within a naïve elastic net setting. Full block Gibbs samplers are provided for implementing the new estimators. The Bayesian graphical ridge and naïve elastic net priors are extended to allow for flexible shrinkage of the off-diagonal elements of the precision matrix. Simulations and practical case studies show that the proposed estimators compare favourably with competing methods and enrich methodological flexibility for data analysis. To this end, a Bayesian approach for estimating differential networks (DN), using the Bayesian adaptive graphical lasso, is introduced. Comparisons to state-of-the-art frequentist techniques highlight the utility of the proposed technique. The novel samplers considered are available in the 'baygel' R package to facilitate usage and exploration for practitioners.

Keywords: Bayesian shrinkage estimation, block Gibbs sampler, differential network, Gaussian graphical model, precision matrix

© University of Pretoria

© University of Pretoria

Table of contents

List of figures xiii

List of tables xvii

Nomenclature xxi

1 Introduction 1
- 1.1 Preliminaries . . . 3
  - 1.1.1 Graph theory . . . 3
  - 1.1.2 Markov random field . . . 4
  - 1.1.3 Gaussian graphical models . . . 4
- 1.2 Related literature . . . 5
- 1.3 Motivation . . . 8
- 1.4 Objectives . . . 8
- 1.5 Contributions . . . 8
- 1.6 Dissertation outline . . . 9

2 The Bayesian Graphical Lasso: Deep Dive 11
- 2.1 The Bayesian graphical lasso . . . 13
  - 2.1.1 Scale mixture of Gaussians . . . 15
- 2.2 Bayesian structure learning . . . 17
- 2.3 Developing a block Gibbs sampler . . . 19
- 2.4 Adaptive extension . . . 22
- 2.5 Conclusion . . . 23

3 Bayesian Graphical Modeling with a Ridge Prior 25
- 3.1 The Bayesian graphical ridge prior . . . 27
- 3.2 Developing a block Gibbs sampler . . . 29
  - 3.2.1 Computational insights . . . 31
- 3.3 Bayesian inference on the precision matrix . . . 32
- 3.4 Bayesian adaptive graphical ridge . . . 35
- 3.5 Synthetic data analysis . . . 36
- 3.6 Bayesian structure learning . . . 41
  - 3.6.1 Cell signalling analysis . . . 45
- 3.7 Conclusion . . . 47

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

4 Bayesian Graphical Modeling with a Naïve Elastic Net Prior **49**
4.1 The naïve Bayesian graphical elastic net prior . . . 51
4.2 Developing a block Gibbs sampler . . . 53
4.2.1 Computational insights . . . 55
4.3 Naïve Bayesian graphical adaptive elastic net . . . 55
4.4 Synthetic data analysis . . . 57
4.5 SEER breast cancer . . . 64
4.6 Conclusion . . . 66

5 Differential Network Analysis: A Bayesian Perspective **67**
5.1 Related works in differential network estimation . . . 69
5.2 The Bayesian differential network . . . 71
5.2.1 Structure learning . . . 72
5.3 Synthetic data analysis . . . 73
5.4 Real data analysis . . . 82
5.4.1 Spam data . . . 82
5.4.2 South African COVID-19 data . . . 83
5.5 Discussion . . . 86

6 Conclusions **87**
6.1 Summary of conclusions . . . 89
6.2 Future work . . . 91

References **93**

Appendix A The Bayesian Graphical Lasso: Deep Dive **99**
A.1 The Bayesian graphical lasso . . . 99
A.1.1 Block Gibbs sampler . . . 99

Appendix B Bayesian Graphical Modeling with a Ridge Prior **101**
B.1 Block Gibbs sampler . . . 101
B.2 Bayesian precision matrix inference . . . 102
B.3 Bayesian structure learning . . . 105

Appendix C Bayesian Graphical Modeling with a Naïve Elastic Net Prior **111**
C.1 SEER breast cancer . . . 111
C.1.1 SEER breast cancer metadata . . . 112

Appendix D Differential Network Analysis: A Bayesian Perspective **117**
D.1 Structure learning . . . 117

Appendix E The baygel Package **121**

Appendix F Derived Publications **131**

© University of Pretoria

List of figures

1.1 *Chapter 1 visual outline.* . . . 3

2.1 *Chapter 2 visual outline.* . . . 13
2.2 *Distribution of the posterior median for a diagonal (a), an off-diagonal (b), partial correlation (c) entry of* $\mathbf{\Omega}$ *for* $p \in \{5, 10, 25, 50, 75\}$ *and the sample size n ranging from 5 to 500.* The size of dots are proportional to the standard error (se) of the elements and the red line indicates the true value. Credible intervals for the partial correlation entry are provided in (d). The red horizontal dashed line indicates the true value and the true precision matrix is set to the identity matrix. . . . 18

3.1 *Chapter 3 visual outline.* . . . 27
3.2 *Coefficient paths for first row of elements in varying* $\mathbf{\Omega}$ *topologies.* Each panel displays the convergence path for the Bayesian graphical lasso (dashed) and the Bayesian graphical ridge (dashed) for $n > p$ and varying σ and $1/\lambda$. The horizontal grey lines (dot-dash) indicate the true values of ω_{ij}. . . . 31
3.3 *Bayesian graphical ridge block Gibbs sampler computational insights.* The amount of time, in minutes, required for 1000 iterations from the Bayesian graphical ridge block Gibbs sampler. . 32
3.4 *Bayesian graphical ridge block Gibbs sampler convergence diagnostics.* The Bayesian inefficiency factor and the Gelman-Rubin diagnostic across multiple network topologies and p/n ratios are displayed in panel (a) and (b) respectively. . . . 33
3.5 *Ratio of Stein's loss across a varying penalty parameter for the full network topology*. All panels display, for the full topology, the ratio of Stein's loss across varying σ. The colour gradient highlights the magnitude of the uncertainty associated with ω_{12} for $p/n \in (0.5, 1)$. The shape of the estimate depicts the percentage error associated with the Bayesian graphical ridge estimates. 34
3.6 *Undirected cell-signalling graphs estimated using the BAGR estimator and various edge selection techniques.* . . . 46

4.1 *Chapter 4 visual outline.* . . . 51
4.2 *Loss comparison between the BAGL, BAGR, BAGENII, BGHS and GELNET on the SEER breast cancer data.* The left-hand panel illustrates Stein's loss on the cohort of patients who survived the disease and the right-hand panel depicts Stein's loss for the the cohort of patients who succumbed to the disease. . . . 64
4.3 *Receiver operating characteristic curves for the BAGL, BAGR, BAGENII, BGHS and GELNET on the SEER breast cancer data for surviving patients.* The curves investigate the performance of the estimator's network recovery abilities for various p/n ratios. . . . 65

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

4.4 *Receiver operating characteristic curves for the BAGL, BAGR, BAGENII, BGHS and GELNET on the SEER breast cancer data for deceased patients.* The curves investigate the performance of the estimator's network recovery abilities for various p/n ratios. 65
4.5 *The inferred networks from the BAGL, BAGR, BAGENII, BGHS and GELNET for the SEER breast cancer data for deceased patients.* The Bayesian networks are determined using a 90% symmetric credible interval. 66

5.1 *Chapter 5 visual outline.* 69
5.2 *F1 score for various topologies, thresholding techniques and varying thresholds.* The curves display the network recovery performance according to the F1 score for various graph structures under varying thresholds for each Bayesian sparsity criterion in (5.12) and (5.13) for dimension $p = 10$. The size of the point is proportional to the standard error obtained during simulation. . . 74
5.3 *Receiver operating characteristic curves for various topologies thresholding techniques.* The ROC curves display the network recovery performance for various graph structures under varying thresholds for each Bayesian sparsity criterion in (5.12) and (5.13) for dimension $p = 10$. 75
5.4 *Estimated DN graphical networks.* Comparison of the true DN, B-net and D-net graphical structure estimates for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model and $p = 10$. 81
5.5 *A comparison of the D-net and B-net DN estimates for the spam emails dataset.* 82
5.6 *South African COVID-19 daily metrics over time.* 7-day moving average filled area line plots with standardised counts for daily new cases; deaths; tests; positive test rate; active cases; tests per active case; recoveries; hospital admissions; hospital discharges; ICU admissions and ventilated patients. 83
5.7 *Bayesian DN estimates of South African COVID-19 data.* The Bayesian DN and corresponding BAGL graphical models between the first two waves; the first wave and the following plateau and finally the difference between the first and second post wave plateaus. The $p-values$ from the Box's M-test for homogeneity of covariance matrices between the contributing precision matrices were all less than 0.001 (Box, 1949). 85

6.1 Chapter 6 visual outline. 89

B.1 *Coefficient paths for first row of elements in varying* $\boldsymbol{\Omega}$ *topologies.* Each panel displays the convergence path for the Bayesian graphical lasso (dashed), Bayesian graphical ridge type I (dashed) and Bayesian graphical ridge type II (dotted) for $n > p$ and varying σ. The horizontal grey lines (dot-dash) indicate the true values of ω_{ij}. 101
B.2 *Coefficient paths for first row of elements in varying* $\boldsymbol{\Omega}$ *topologies.* Each panel displays the convergence path for the Bayesian graphical lasso (dashed), Bayesian graphical ridge type I (dashed) and Bayesian graphical ridge type II (dotted) for $n > p$ and varying σ. The horizontal grey lines (dot-dash) indicate the true values of ω_{ij}. 102
B.3 *Ratio of Stein's loss across a varying penalty parameter for the AR(1) network topology.* 102
B.4 *Ratio of Stein's loss across a varying penalty parameter for the cluster network topology.* 102
B.5 *Ratio of Stein's loss across a varying penalty parameter for the star network topology.* 103
B.6 *Ratio of Stein's loss across a varying penalty parameter for the scale-free network topology.* . . . 103
B.7 *Ratio of Stein's loss across a varying penalty parameter for the hub network topology.* 103
B.8 *Ratio of Stein's loss across a varying penalty parameter for the full network topology.* 104

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

B.9 *Ratio of Stein's loss across a varying penalty parameter for the band network topology.* 104
B.10 *Ratio of Stein's loss across a varying penalty parameter for the AR(2) network topology.* 104
B.11 *Ratio of Stein's loss across a varying penalty parameter for the circle network topology.* 105

C.1 *Loss comparison between the BAGL, BAGR, BAGENII, BGHS and GELNET on the SEER breast cancer data.* The left-hand panel illustrates the Frobenius loss (referred to as L2 loss on the y-axis) on the cohort of patients who survived the disease and the right-hand panel depicts Frobenius loss for the the cohort of patients who succumbed to the disease. 111
C.2 *The inferred networks from the BAGL, BAGR, BAGENII, BGHS and GELNET for the SEER breast cancer data for alive patients.* The Bayesian networks are determined using a 90% symmetric credible interval. 112

D.1 *F1 score for various topologies, thresholding techniques and varying thresholds.* The curves illustrate how varying thresholds and Bayesian sparsity criteria in (5.12) and (5.13) affect the F1 score of network recovery for graphs with a dimension of $p = 30$, with point sizes reflecting standard error magnitude. 117
D.2 *F1 score for various topologies, thresholding techniques and varying thresholds.* The curves illustrate how varying thresholds and Bayesian sparsity criteria in (5.12) and (5.13) affect the F1 score of network recovery for graphs with a dimension of $p = 100$, with point sizes reflecting standard error magnitude. 117
D.3 *Receiver operating characteristic curves for various topologies thresholding techniques.* The ROC curves display the network recovery performance for various graph structures under varying thresholds for each Bayesian sparsity criterion in (5.12) and (5.13) for dimension $p = 30$. 118
D.4 *Receiver operating characteristic curves for various topologies thresholding techniques.* The ROC curves display the network recovery performance for various graph structures under varying thresholds for each Bayesian sparsity criterion in (5.12) and (5.13) for dimension $p = 100$. 119

© University of Pretoria

© University of Pretoria

List of tables

3.1 *Summary of the precision matrix algorithms considered in the synthetic data analysis study.* . . . 36
3.2 *Summary of the F norm and Stein's loss across multiple network topologies and estimators for* $p/n = 0.5$. The median loss values and (standard errors) reported here are based on 30 replications for the BAGL, BAGR, ALTI and ALTII estimators. The Bayesian estimates are obtained as the mean of the posterior distribution. The cell colour of the best performing values are set to grey. . 38
3.3 *Summary of the F norm and Stein's loss across multiple network topologies and estimators for* $p/n = 1$. The median loss values and (standard errors) reported here are based on 30 replications for the BAGL, BAGR, ALTI and ALTII estimators. The Bayesian estimates are obtained as the mean of the posterior distribution. The cell colour of the best performing values are set to grey. . 39
3.4 *Summary of the F norm and Stein's loss across multiple network topologies and estimators for* $p/n = 10$. The median loss values and (standard errors) reported here are based on 30 replications for the BAGL, BAGR, ALTI and ALTII estimators. The Bayesian estimates are obtained as the mean of the posterior distribution. The cell colour of the best performing values are set to grey. . 40
3.5 *Summary of the heuristic structure learning techniques considered in the synthetic data analysis study.* . . . 41
3.6 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 0.5$. The median loss values reported here are based on 30 replications for the BAGL estimator. The cell colour of the best performer is set to grey. 43
3.7 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 0.5$. The median loss values reported here are based on 30 replications for the BAGR estimator. The cell colour of the best performer is set to grey. 44
3.8 *Evaluation of established molecule edges by BAGR and BAGL. Refer to Table 3 in (Sachs et al., 2005).* . . . 46

4.1 *Summary of the precision matrix algorithms considered in the synthetic data analysis study.* . . . 58
4.2 *Summary of network signal strength and structure recovery performance across multiple network topologies and estimators for* $p/n = 0.5$. The median performance values and (standard errors) reported here are based on 30 replications for the BAGENI, BAGENII, BAGL, BAGR, BGHS and GELNET estimators. The best performance value is boldfaced. Cell colours correspond only to the BAGL, BAGR, BAGENI and BAGENII signal strength metrics, with darker colours indicating the best performance value and lighter colours representing the runner-up. . . . 60

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

4.3 *Summary of network signal strength and structure recovery performance across multiple network topologies and estimators for* $p/n = 1$. The median performance values and (standard errors) reported here are based on 30 replications for the BAGENI, BAGENII, BAGL, BAGR, BGHS and GELNET estimators. The best performance value is boldfaced. Cell colours correspond only to the BAGL, BAGR, BAGENI and BAGENII signal strength metrics, with darker colours indicating the best performance value and lighter colours representing the runner-up. 61

4.4 *Summary of network signal strength and structure recovery performance across multiple network topologies and estimators for* $p/n = 4$. The median performance values and (standard errors) reported here are based on 30 replications for the BAGENI, BAGENII, BAGL, BAGR, BGHS and GELNET estimators. The best performance value is boldfaced. Cell colours correspond only to the BAGL, BAGR, BAGENI and BAGENII signal strength metrics, with darker colours indicating the best performance value and lighter colours representing the runner-up. 62

4.5 *Summary of network signal strength and structure recovery performance across multiple network topologies and estimators for* $p/n = 10$. The median performance values and (standard errors) reported here are based on 30 replications for the BAGENI, BAGENII, BAGL, BAGR, BGHS and GELNET estimators. The best performance value is boldfaced. Cell colours correspond only to the BAGL, BAGR, BAGENI and BAGENII signal strength metrics, with darker colours indicating the best performance value and lighter colours representing the runner-up. 63

5.1 *Synthetic study median loss results.* Summary of F norm, EL1, EL2, MAXEL1 and MINEL1 for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model. The median loss values reported here are based on 40 replications for both the B-net and D-net estimators. The best performing values are boldfaced. 78

5.2 *Synthetic study standard error loss results.* Summary of F norm, EL1, EL2, MAXEL1 and MINEL1 for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model. The median standard errors reported here are based on 40 replications for both the B-net and D-net estimators. The best performing values are boldfaced. 79

5.3 *Synthetic study median performance results.* Summary of SE, SP, MCC, F1 and FNR for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model. The median performance scores reported here are based on 40 replications for both the B-net and D-net estimators. The best performing values are boldfaced. 80

B.1 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 1$. The median loss values reported here are based on 30 replications for the BAGR estimator. The cell colour of the best performer is set to grey. 106

B.2 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 1$. The median loss values reported here are based on 30 replications for the BAGL estimator. The cell colour of the best performer is set to grey. 107

B.3 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 1$. The median loss values reported here are based on 30 replications for the BAGL estimator. The cell colour of the best performer is set to grey. 107

B.4 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 10$. The median loss values reported here are based on 30 replications for the BAGR estimator. The cell colour of the best performer is set to grey. 108

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

B.5 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 10$. The median loss values reported here are based on 30 replications for the BAGL estimator. The cell colour of the best performer is set to grey. 109

© University of Pretoria

© University of Pretoria

Nomenclature

Greek Symbols

λ	Shrinkage parameter of the Bayesian (adaptive) graphical lasso
$\boldsymbol{\mu}$	Population mean vector for a multivariate Gaussian probability density function
$\boldsymbol{\Omega}$	Precision matrix
ω_{ij}	Element ij of the precision matrix $\boldsymbol{\Omega}$
$\mathbf{P}$	Partial correlation matrix
ϕ	An empty set
σ	Shrinkage parameter of the Bayesian (adaptive) graphical ridge
$\boldsymbol{\Sigma}$	Covariance matrix

Other Symbols

$\mathbf{A}$	Adjacency matrix
$C^{-1}_{\{\eta_{ij},\lambda_{ij}\}_{i<j}}$	Normalising constant of the type II naïve Bayesian adaptive graphical elastic net prior for $\boldsymbol{\Omega}$
$C^{-1}_{\{\kappa_{ij}\}_{i<j}}$	Normalising constant of the Bayesian adaptive graphical ridge prior for $\boldsymbol{\Omega}$
$C^{-1}_{\{\kappa_{ij},\lambda_{ij}\}_{i<j}}$	Normalising constant of the type I naïve Bayesian adaptive graphical elastic net prior for $\boldsymbol{\Omega}$
$C^{-1}_{\{\lambda_{ij}\}_{i\leq j}}$	Normalising constant of the Bayesian adaptive graphical lasso prior for $\boldsymbol{\Omega}$
$\text{DE}(\cdot)$	Double exponential probability density function notation
$\det(\cdot)$	Determinant of a matrix
$\boldsymbol{\Delta}$	Differential network
$\text{EXP}(\cdot)$	Exponential probability density function notation
$\mathbb{E}(\cdot)$	Expected value operator
$G(V,E)$	Graph with nodes V and edges E
$\mathbf{I}_p$	Identity matrix of dimension p

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$\mathrm{IG}(\cdot)$ Inverse Gaussian probability density function notation

$\mathrm{IGA}(\cdot)$ Inverse gamma probability density function notation

$\mathbb{1}$ Indicator function where its value equals 1 if the condition is met and 0 otherwise

ℓ_1 Absolute value penalty

$\|\mathbf{A}\|_F$ Frobenius norm of an arbitrary matrix $\mathbf{A}$

$\|\mathbf{A}\|_1$ ℓ_1 norm of an arbitrary matrix $\mathbf{A}$

$\mathbb{M}^+$ Space of positive definite matrices

$\mathrm{N}_p(\cdot)$ p-variate Gaussian probability density function notation

$\mathrm{N}(\cdot)$ Gaussian probability density function notation

n Sample size

$\mathbb{R}^p$ Real coordinate space of dimension p

$\mathbf{S}$ Sample covariance matrix

$\mathrm{TN}(\cdot)$ Truncated Gaussian probability density function notation

$\mathrm{trace}(\cdot)$ Trace of a matrix

$\mathrm{W}(\cdot)$ Wishart probability density function notation

Acronyms / Abbreviations

ADDM Alternating direction method of multipliers

ALTI Frequentist type I graphical ridge

ALTII Frequentist type II graphical ridge

AR(1) First-order autoregressive process

AR(2) Second-order autoregressive process

B-net Bayesian differential network

BAGENI Type I naïve Bayesian adaptive graphical elastic net

BAGENII Type II naïve Bayesian adaptive graphical elastic net

BAGL Bayesian adaptive graphical lasso

BGL Bayesian graphical lasso

BAGR Bayesian adaptive graphical ridge

BGR Bayesian graphical ridge

BGHS Bayesian graphical horseshoe

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

BIC Bayesian information criterion

CLIME Constrained ℓ_1 minimisation estimator

D-net Iterative shrinkage-thresholding differential network

DN Differential network

FNR False negative rate

FP False positive

GCL Graphical cooperative lasso

GELNET Frequentist graphical elastic net

GGM Gaussian graphical model

GMRF Gaussian Markov random field

IGL Intertwined graphical lasso

iid Independent and identically distributed

MAP Maximum A Posteriori

MAXEL1 ℓ_1 loss on the largest eigenvalue

FN Matthews correlation coefficient

MCMC Markov chain Monte Carlo

EL1 ℓ_1 eigenvalue loss

EL2 Frobenius eigenvalue loss

MINEL2 ℓ_1 loss on the smallest eigenvalue

MOSS Mode oriented stochastic search

MPL Marginal pseudo-likelihood

MRF Markov random field

ROC Receiver operating characteristic

SCAD Smoothly clipped absolute deviation

SEER Surveillance, epidemiology, and end results

TAII Absolute value of the decision-theoretic Wishart thresholding criterion

TAI Decision-theoretic Wishart thresholding criterion

TN True negative

TP True positive

© University of Pretoria

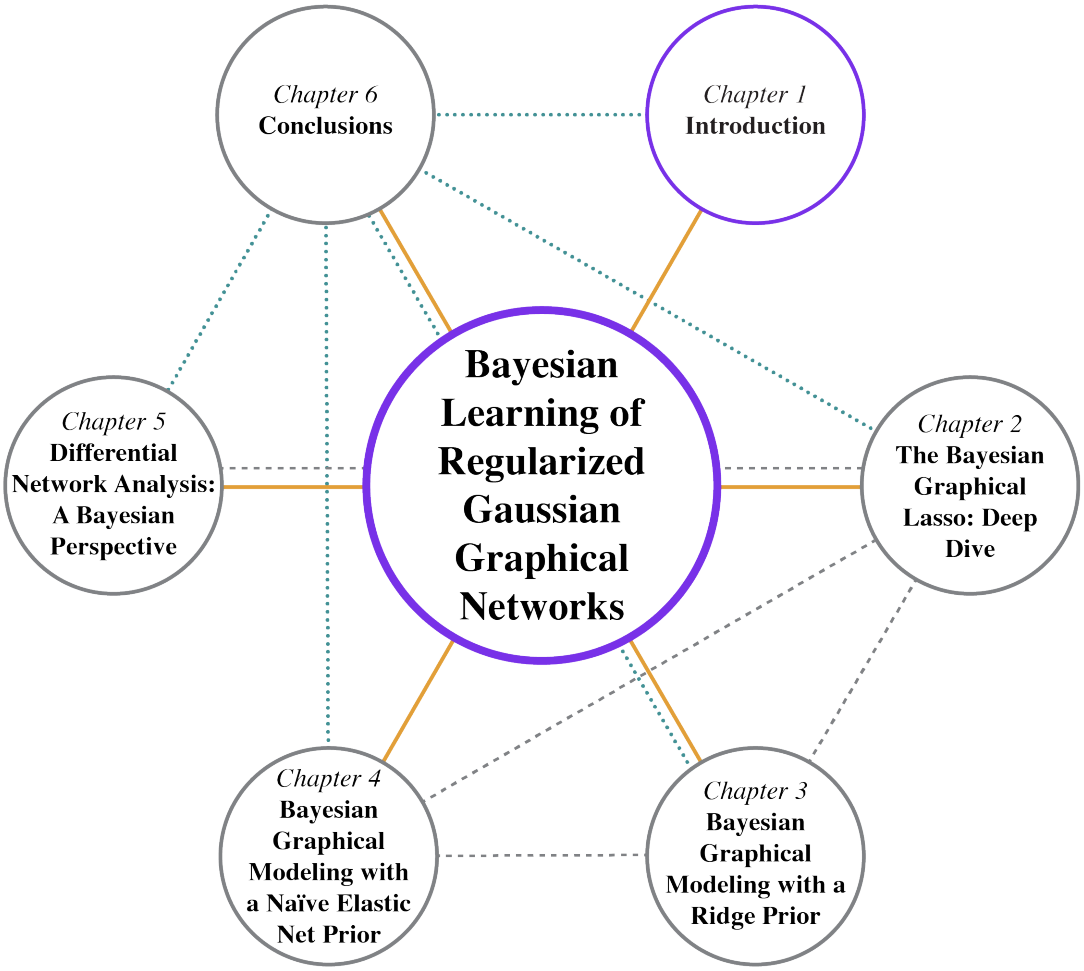
© University of Pretoria

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chapter 1

Introduction



© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

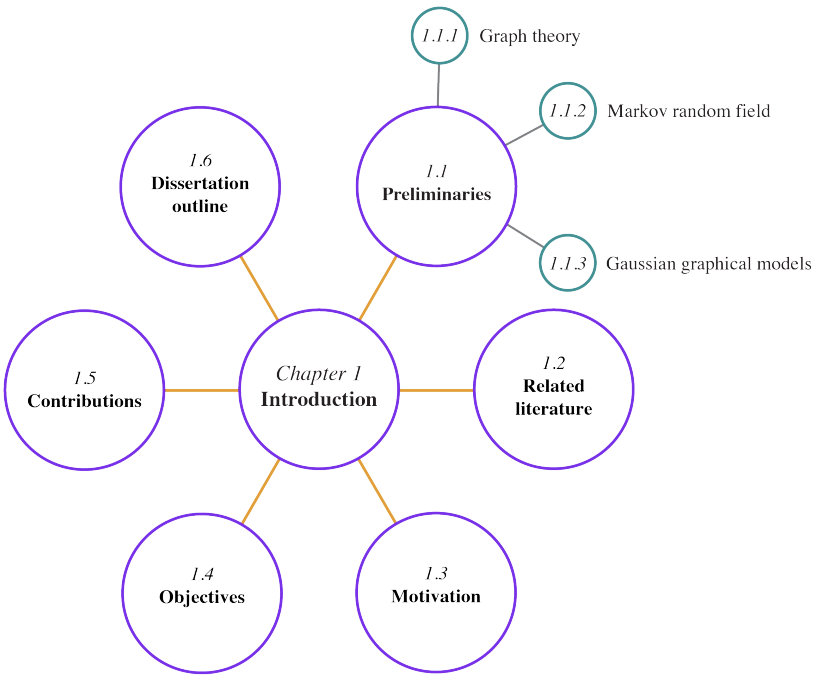


Fig. 1.1 *Chapter 1 visual outline.*

In a data-rich landscape, where information spans a wide range of dimensions across various fields of research, graphical models play a pivotal role. The works authored by (Whittaker, 1990), (Lauritzen, 1996) and (Edwards, 2000) offer a thorough exploration of the statistical theory and applications underpinning graphical models. Combined, these works serve as invaluable resources for those seeking a deep understanding of graphical models. That being said, the field remains a vibrant and actively evolving area of research, for example the recent work by (Chen, 2023) offers an up-to-date review of the latest advancements, covering both theoretical developments and computational methodologies. In essence, graphical modelling is a branch of multivariate analysis where *graphs* are used to illustrate both associational and causal dependencies among model variables. Similar to many statistical fields, the domain of graphical models is multifaceted, and this dissertation focuses on a specific area within it, namely *undirected* graphs for multivariate Gaussian models from a Bayesian viewpoint. The previous statement introduces intricate concepts which require clarification for a smoother reading experience. Figure 1.1 provides a visual outline of this chapter.

1.1 Preliminaries

This section is devoted to basic notions and concepts required for the introduction of Gaussian graphical models (GGM).

1.1.1 Graph theory

First, by adopting the notation in Chapter 2 of (Lauritzen, 1996), the conditional (in)dependencies between pairs of variables can be represented as a (simple finite) graph $G = (V, E)$, where $V = \{1, \cdots, p\}$ is a finite

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

set of *vertices* (also referred to as nodes). Each vertex corresponds to a random variable X_i. The edge set $E = \{(i, j) \in V \times V : i \neq j\}$, i.e. ordered pairs of distinct vertices, corresponds to conditionally dependent variables. Next, let $E^- = \{(i, j) \in V \times V : i \neq j\ , (i, j) \notin E\}$ define the set of edges not present in G. It follows that edges $(i, j) \in E$ with both (i, j) and (j, i) in E are referred to as undirected, whereas if one of the pairs is not in E, the edges are directed. Additional concepts and notions are available in (Baz et al., 2023).

1.1.2 Markov random field

Second, the bridge between the basic graph theory and probabilistic graphical models is considered. The concept of a Markov random field (MRF) links the the conditional (in)dependence structure of a vector of random variables to the finite graph introduced before. That is, for the random vector of dimension p, $\bar{X} = \{X_1, X_2, \cdots, X_p\}$, the MRF consists of an undirected graph $G = (V, E)$ with $V = \{1, \cdots, p\}$, $E \subseteq V \times V$, and the Markov properties of interest are given by:

- The pairwise Markov property: $X_i \perp X_j \mid \bar{X}_{-\{ij\}}$. Here, nodes X_i and X_j are conditionally independent given all other nodes $\bar{X}_{-\{ij\}}$. The symbol $\perp$ denotes the independence between two nodes and $\bar{X}_{-\{ij\}}$ denotes all nodes in V except for nodes X_i and X_j.
- The local Markov property: $X_i \perp \bar{X}_{\{V-i\}} \mid \bar{X}_{N_i}$. Here, any node X_i, where $i \in V$, is conditionally independent of any other node in V given all of node X_i's neighbours - N_i. The symbol $\bar{X}_{\{V-i\}}$ denotes all nodes in V except for node X_i and $\bar{X}_{N_i}$ the nodes which are direct neighbours to node X_i, i.e. there is an edge from X_i to each node in $\bar{X}_{N_i}$.
- The global Markov property: $\bar{X}_{V_1} \perp \bar{X}_{V_2} \mid \bar{X}_{V_3}$. For any pairwisely disjoint $V_1, V_2, V_3 \subseteq V$ with $V_1, V_2 \neq \phi$ and such that V_3 separates V_1 and V_2. In other words, if any path from a node in V_1 to a node in V_2 includes at least one node in V_3. The symbol ϕ denotes the empty set.

If the random vector $\bar{X} = \{X_1, X_2, \cdots, X_p\}$ is distributed according to the multivariate Gaussian distribution, then all three properties above are equivalent. In addition to this, if any of the three properties above are satisfied, $\bar{X}$ is termed a Gaussian random Markov Field (GMRF) over G. A rigorous treatment of the concepts and notions are available in Chapter 3 of (Lauritzen, 1996) and further insights in (Baz et al., 2023).

1.1.3 Gaussian graphical models

Lastly, regarding the distributional assumption going forward; let $\bar{X} = (X_1, X_2, \cdots, X_p) \sim N_p(\mu, \mathbf{\Sigma})$, where $\mu \in \mathbb{R}^p$, $\mathbf{\Sigma} \in \mathbb{M}^+$ and $\mathbb{M}^+$ denotes the space of positive definite matrices (Uhler, 2017). In this setting, any two variables X_i and X_j are conditionally independent if, given all other variables $\bar{X}_{-\{ij\}}$, the (i, j)'th entry of the inverse covariance matrix (commonly referred to as the precision or concentration matrix) $\mathbf{\Omega} = (\omega_{ij}) = \mathbf{\Sigma}^{-1}$ is zero. More formally from (Speed and Kiiveri, 1986), Chapter 5 of (Lauritzen, 1996), Chapter 6 of (Whittaker, 1990) and (Baz et al., 2023),

$$X_i \perp X_j \mid \bar{X}_{-\{ij\}} \Leftrightarrow \omega_{ij} = \omega_{ji} = 0. \tag{1.1}$$

which is equivalent to the following; let $G = (V, E)$ with $V = \{1, \cdots, p\}$, $E \subseteq V \times V$ and $\bar{X}$ as above, then $\bar{X}$ is a GMRF over G if,

$$\{i, j\} \notin E \Leftrightarrow \omega_{ij} = \omega_{ji} = 0.$$

© University of Pretoria

The implication of these results imply that the graph G is fully characterised by $\boldsymbol{\Omega}$ and was first introduced as a covariance selection problem statement in (Dempster, 1972). The partial correlation matrix $\mathbf{P} = (\rho_{ij})$ is also typically used to characterise G, where the partial correlation between variables X_i and X_j is given by ρ_{ij}, and

$$\rho_{ij} = \frac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}.$$

Estimating $\boldsymbol{\Omega}$ and obtaining G under the given distributional assumptions is referred to as Gaussian graphical modelling. Here, G is commonly represented by the *adjacency* matrix $\mathbf{A} = (a_{ij})$ such that $a_{ij} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$, otherwise. Finally the assumption going forward is as follows: without loss of generality let $\mathbf{Y} = (\bar{y}_1, \bar{y}_2, ..., \bar{y}_n)$ be the matrix of observations where each $\bar{Y}_i \sim N\left(\mathbf{0}, \boldsymbol{\Omega}^{-1}\right)$ is *independent and identically distributed* (iid) and the GGM is fully characterised by the parameter set of $\boldsymbol{\Omega}$ (Atay-Kayis and Massam, 2005). A note going forward; the notation used to represent vectors in Chapters 2 - 5 will make use of bold facing. The use of the "bar" notation was necessary within the preliminaries in order to distinguish between matrices and vectors.

1.2 Related literature

The body of knowledge is rich with literature that addresses the estimation of GGMs from a frequentist perspective. Some foundational work in this field includes the nodewise regression approach by (Meinshausen and Bühlmann, 2006). Here, a single variable is chosen as the response whilst the remaining variables are treated as predictors, this process is repeated for each variable, with penalisation induced via the lasso (Tibshirani, 1996). In this approach, the zero entries in the estimated lasso regression coefficients are used as an estimate of G, however, a full estimate of $\boldsymbol{\Omega}$ is not provided. This issue is addressed in (Zhou et al., 2011), who utilise an unpenalised maximum likelihood estimator to estimate $\boldsymbol{\Omega}$ using the graph obtained from the nodewise regression approach before. Penalised likelihood framework techniques are a popular method of choice due to their ability to simultaneously estimate the precision matrix and infer the associated graph. For example, (Yuan and Lin, 2007) simultaneously estimate $\boldsymbol{\Omega}$ and its corresponding structure by setting elements of estimated precision matrix to exactly zero. This is achieved through the maximisation of an ℓ_1 type penalised Gaussian log-likelihood for positive definite matrices. Similar approaches are also considered in (Banerjee et al., 2008), (Yuan, 2008), the seminal graphical lasso in (Friedman et al., 2008), the smoothly clipped absolute deviation (SCAD) in (Fan et al., 2009) and the constrained ℓ_1 minimisation estimator (CLIME) in (Cai et al., 2011), to name but a few. An ℓ_2 type penalised likelihood approach for the estimation of $\boldsymbol{\Omega}$ is considered in (Wieringen and Peeters, 2016) and (Kuismin et al., 2017). The latter authors refer to their estimator as the ridge type operator for precision matrix estimation (ROPE) for various target matrices. The former labels it the alternative ridge precision estimator, with the type I referring to the case of a non-zero positive definite target matrix and type II for a null target matrix. Here, the estimates of $\boldsymbol{\Omega}$ do not contain exact zeros and a heuristic structure learning procedure is required to recover G. A generalised ridge estimator which allows for element-wise penalisation is described in (van Wieringen, 2019). Finally, inspired by the elastic net regression in (Zou and Hastie, 2005), and the formulation of ridge estimators above, (Kovács et al., 2021) introduces the graphical elastic which combines both the ℓ_1 and ℓ_2 penalties within the Gaussian log-likelihood based precision matrix estimation and structure determination.

Under the Bayesian lens, there has been a rapid increase of research dedicated towards the estimation of GGMs. The strength of Bayesian methods lies in managing model uncertainty through posterior distributions and prior knowledge (Williams et al., 2018). Techniques in this domain can broadly be categorised into three types based on the parameter space they investigate: joint space of graphs and precision matrices, space of graphs, and space of precision matrices (Vogels et al., 2023).

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

In the GGM paradigm, Bayesian computation of the posterior graph distribution, conditional on the data, becomes exponentially impractical with more nodes due to complex integrals (Mohammadi et al., 2023b). Resultantly, the most popular approaches estimate the joint posterior distribution of the graph and precision matrix ((Peterson et al., 2015), (van den Boom et al., 2022a), (Mohammadi et al., 2023a)) through means of efficient space exploration. Despite its popularity, this approach is computationally complex because of the large joint space of graphs and precision matrices. Therefore, most Bayesian methods use Markov chain Monte Carlo (MCMC) sampling to explore this space. Noteworthy contributions within this domain include the reversible jump MCMC algorithm by (Green, 1995) which is extended to the GGM case by (Dobra et al., 2011). The conjugate G-Wishart is commonly used as the prior distribution of the precision matrix conditioned the graph ((Dawid and Lauritzen, 1993), (Roverato, 2002), (Letac and Massam, 2007)), its popularity stems from its edge selection capabilities. Methods aimed at reducing the computational demand within this approach, notably due to the ratio of prior normalising constants of the G-Wishart (Mohammadi et al., 2023a), have gained considerable attention. Computationally efficient sampling algorithms includes the works by ((Wang and Li, 2012), (Cheng and Lenkoski, 2012), (Lenkoski, 2013), (van den Boom et al., 2022a)). Further efficiency is achieved via the birth-death MCMC algorithm by (Mohammadi and Wit, 2015) and the block Gibbs algorithm using a spike-and-slab prior distribution is presented in (Wang, 2015).

Bayesian methods which only focus on the posterior distribution of the graph, bypass the computational costs and complexity of the normalising constants for the G-Wishart distribution. As before, techniques rooted in this methodology are well established literature. For example, an efficient Monte Carlo method for computing the normalising constants for the G-Wishart distribution is detailed in (Atay-Kayis and Massam, 2005). Similarly, (Lenkoski and Dobra, 2011) use a Laplace approximation in their adaption of the mode oriented stochastic search (MOSS) algorithm to find graphs with a high posterior probability. Developed within the literature of unbiased estimation in state-space models, (van den Boom et al., 2022b) propose an efficient unbiased MCMC algorithm based on coupling of Markov chains for a broad class of posterior distributions. Next, (Mohammadi et al., 2023b) introduce two efficient MCMC algorithms which bypass the normalising constants for the G-Wishart distribution by replacing the Gaussian likelihood function with a product of conditional likelihood functions i.e. marginal pseudo-likelihood (MPL) algorithms (Leppä-Aho et al., 2017).

Bayesian techniques focusing on the posterior distribution of the precision matrix remove the necessity of estimating the graph within the sampling procedures. These techniques typically make use of absolute continuous priors for $\mathbf{\Omega}$. This streamlined approach is responsible for the recent increase popularity. These techniques typically benefit from reduced computational complexity when compared to algorithms for the joint space, however, here zero prior belief is placed on G so most methods here require heuristic treatments for structure learning. The first technique (and arguably the catalyst and point of reference for the subsequent techniques) is the seminal Bayesian graphical lasso by (Wang, 2012). The prior on $\mathbf{\Omega}$ is specified as the product of double exponential distributions for the off-diagonal elements and a exponential distributions for the main diagonal elements. Similar to the Bayesian lasso (Park and Casella, 2008), the posterior mode of the Bayesian graphical lasso coincides with the graphical lasso. Sampling is achieved through a block Gibbs sampler and structure learning through a thresholding treatment of the estimated posterior partial correlation matrix. The same prior is presented almost simultaneously in (Khondker et al., 2013), however, here sampling is achieved through a Metropolis-Hasting algorithm and structure learning via a treatment of the credible intervals on the posterior distribution of $\mathbf{\Omega}$. Interestingly, building upon the Bayesian graphical lasso, (Banerjee and Ghosal, 2015) propose a mixture prior which is capable of estimating both $\mathbf{\Omega}$ and G. The components of this mixture include the Bayesian graphical lasso prior for $\mathbf{\Omega}$ to ensure positive definiteness and a point-mass prior on G. (Kundu et al., 2019) propose shrinkage priors in the form of mixtures of inverse–Wishart priors for $\mathbf{\Sigma}$ where the conjugacy of their formulation allows for simulation of

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$\boldsymbol{\Sigma}$. Structure learning is achieved heuristically through a sophisticated decision theoretic approach to infer absent edges within penalised credible intervals of $\boldsymbol{\Omega}$. Furthermore, (Li et al., 2019) introduce the Bayesian graphical horseshoe estimator which places horseshoe, (Carvalho et al., 2010), priors on the off-diagonal elements of $\boldsymbol{\Omega}$, and a non-informative prior on the diagonal elements. The estimator operates under the assumption of an arbitrary sparsity pattern for $\boldsymbol{\Omega}$ and performs exceptionally well in sparse high-dimensional problems. Similarly, (Sagar et al., 2021) construct their prior in a similar fashion to the latter, except here independent horseshoe-like priors, (Bhadra et al., 2021), are used for the off-diagonal elements of $\boldsymbol{\Omega}$. The prior here is available in closed form which allows the authors to infer theoretical properties on posterior contraction rates which are not available under the former. Structure learning for both horseshoe based priors is achieved through the use of credible intervals. Despite the rich literature on Bayesian techniques within the space of precision matrix estimation, a Bayesian treatment of ridge estimation under arbitrary sparsity patterns is missing. Such a technique will be beneficial in situations where where sparsity or excessive shrinkage is not required. Next, an elastic net formulation in this Bayesian setting is also not available which will prove helpful in situations where the required sparsity is not known apriori. It is worth mentioning at this point that the paradigm of Bayesian techniques focused on precision matrix estimation under arbitrary sparsity patterns is exactly where the contributions of this dissertation are nested, the details of which are outlined in the subsequent section.

It is clear that GGMs have undeniably drawn significant attention within statistical literature. This popularity is due to their ability in estimating $\boldsymbol{\Omega}$, and the associated graph G, for multivariate Gaussian data with many variables, making them valuable for studying conditional (in)dependencies. GGMs are increasingly applied across scientific disciplines, including biology (Yin and Li, 2011), meteorology (Keune et al., 2014), psychology (Williams et al., 2018), and neuroscience (Yin et al., 2020), to name a few. There is also growing interest in the scenario where networks may undergo changes over time or where they may differ between states. In this setting a *differential network* (DN) is an important tools for capturing the changes in conditional (in)dependencies between two samples. Similar to the preliminaries discussed before, let that the observations, $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{n_1}$ and $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{n_2}$, be generated from multivariate Gaussian distributions, $N_p(\mathbf{0}, \boldsymbol{\Omega}_1^{-1})$ and $N_p(\mathbf{0}, \boldsymbol{\Omega}_2^{-1})$, respectively. Here, n_1 and n_2 represent the respective sample sizes, which need not be equal. The central focus of this analysis lies in the estimation of the DN (difference between precision matrices), denoted as $\boldsymbol{\Delta} = \boldsymbol{\Omega}_2 - \boldsymbol{\Omega}_1$. A comprehensive statistical perspective is proved by (Shojaie, 2020). Techniques for estimating $\boldsymbol{\Delta}$ can largely be classified into two main categories. The first estimating the individual precision matrices, $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ separately; where the estimated DN is the difference between the estimated precision matrices. For example, the methods for GGM estimation outlined above. The second methodology estimates both the precision matrices simultaneously. The approach here, typically penalises a joint loss function for both precision matrices. (Chiquet et al., 2011) provide a methodology for inference and estimation of functions of GGMs. In particular, the 'Intertwined Graphical Lasso' (IGL) approach biases the estimation of the precision matrices towards a common value. More so, their 'Graphical Cooperative Lasso' (GCL) utilises a group-penalty for solutions that favour a common sparsity pattern. (Guo et al., 2011) and (Zhu and Li, 2018) estimate separate graphical models using a joint penalised loss function. (Zhao et al., 2014) propose a method for estimating $\boldsymbol{\Delta}$ directly which relaxes the need for both individual precision matrices to be sparse nor be estimated directly. Similarly, (Yuan et al., 2017) and (Jiang et al., 2018) utilise an alternating direction method of multipliers (ADMM) algorithm for estimating $\boldsymbol{\Delta}$ from their joint ℓ_1 penalised convex loss function. More recently, (Tang et al., 2020) introduce a computationally efficient iterative shrinkage-thresholding algorithm for minimising the ℓ_1 loss function defined in (Jiang et al., 2018). At the time of writing, DN estimation from the Bayesian perspective, remains unexplored in the body of knowledge, and is therefore of also of particular interest in this dissertation, the details of which are outlined below.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

1.3 Motivation

Bayesian shrinkage techniques are attractive for accurate precision matrix estimation for GGMs with low computational cost under moderate dimensions. An example of this is the Bayesian graphical lasso of (Wang, 2012), who mentions that their framework can be extended to other shrinkage related methods for graphical modelling. Moreover, in the concluding remarks of (Wieringen and Peeters, 2016), the authors foresee a Bayesian investigation of their computationally efficient ridge treatment for precision matrices. Next, (Kovács et al., 2021) focus heavily on methodological extensions for flexible data analysis with their elastic net type treatment of precision matrices. Finally, as noted by (Shutta et al., 2023) it is not always straightforward to establish universal guidelines for choosing a GGM estimation method without apriori knowledge.These deductions provide the fuel for enriching Bayesian methodology in the area of Gaussian graphical modelling.

In summary: the flexibility and computational efficiency of the Bayesian graphical lasso can be extended to include Bayesian graphical ridge and elastic net formulations for varying sparsity requirements. This enriches the Bayesian toolkit for practitioners of graphical modelling, with methodological extensions and user-friendly, publicly available software. To quote (Williams, 2021): "However, in my view, the greatest strength of Bayesian methods is the richness of information provided from drawing samples from the posterior distribution...".

1.4 Objectives

The objectives of this dissertation are:

- Thoroughly review the technical constructs of the Bayesian graphical lasso.
- Propose a ridge-type prior within the framework of absolute continuous priors, arbitrary sparsity patterns of the precision matrix and a multivariate Gaussian data model.
- Propose an elastic net type prior within the framework of absolute continuous priors, arbitrary sparsity patterns of the precision matrix and a multivariate Gaussian data model.
- Formulate a differential network estimate from a Bayesian perspective.

1.5 Contributions

The contributions this dissertation sets out to make in the field of Bayesian Gaussian graphical models are:

- The development of the Bayesian graphical ridge-type prior for precision matrix estimation in the setting of absolute continuous priors, arbitrary sparsity patterns and a multivariate Gaussian data model.
- Extend the Bayesian graphical ridge-type prior to the Bayesian adaptive graphical ridge-type hierarchical prior for flexible shrinkage on the off-diagonal elements of the precision matrix.
- The development of the naïve Bayesian graphical elastic net prior for precision matrix estimation in the setting of absolute continuous priors, arbitrary sparsity patterns and a multivariate Gaussian data model.
- Extend the naïve Bayesian graphical elastic net prior to the naïve Bayesian adaptive graphical elastic net hierarchical prior for flexible shrinkage on the off-diagonal elements of the precision matrix.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

- Provide a full Gibbs sampler for implementing the Bayesian graphical (adaptive) ridge-type and naïve (adaptive) elastic net estimators.
- The presentation of differential network estimation using Bayesian precision matrix estimation techniques.
- Develop an R, (R Core Team, 2023), package implementing the Bayesian (adaptive) graphical lasso; (adaptive) graphical ridge-type and naïve (adaptive) elastic net block Gibbs samplers.

1.6 Dissertation outline

- **Chapter** 2 provides a thorough review of the Bayesian graphical lasso, the critical foundation for the research presented in the subsequent chapters.
- **Chapter** 3 introduces the Bayesian graphical ridge-type prior, the Bayesian adaptive graphical ridge-type hierarchical prior and their corresponding block Gibbs samplers for precision matrix estimation.
- **Chapter** 4 introduces the naïve Bayesian graphical elastic net prior, the naïve Bayesian adaptive graphical elastic net hierarchical prior and their corresponding block Gibbs samplers for precision matrix estimation.
- **Chapter** 5 explores how differential network estimation can be achieved by employing Bayesian techniques for the estimation of its precision matrix components.
- **Chapter** 6 summarises the main findings of this dissertation and the scope for future work.

The set of appendices include:

- **Appendix** A provides technical supplementary material related to the properties of the Bayesian graphical lasso in Chapter 2.
- **Appendix** B provides additional simulation results and properties of the Bayesian graphical ridge-type block Gibbs sampler, introduced in Chapter 3.
- **Appendix** C provides additional simulation results for the naïve Bayesian graphical elastic net block Gibbs sampler, as well as, supplementary material to enrich the application study considered in Chapter 4.
- **Appendix** D provides additional simulation results for the heuristic structure learning techniques considered in Chapter 5.
- **Appendix** E describes the 'baygel' R package for the block Gibbs samplers used in this dissertation.
- **Appendix** F lists the publications derived from this dissertation.

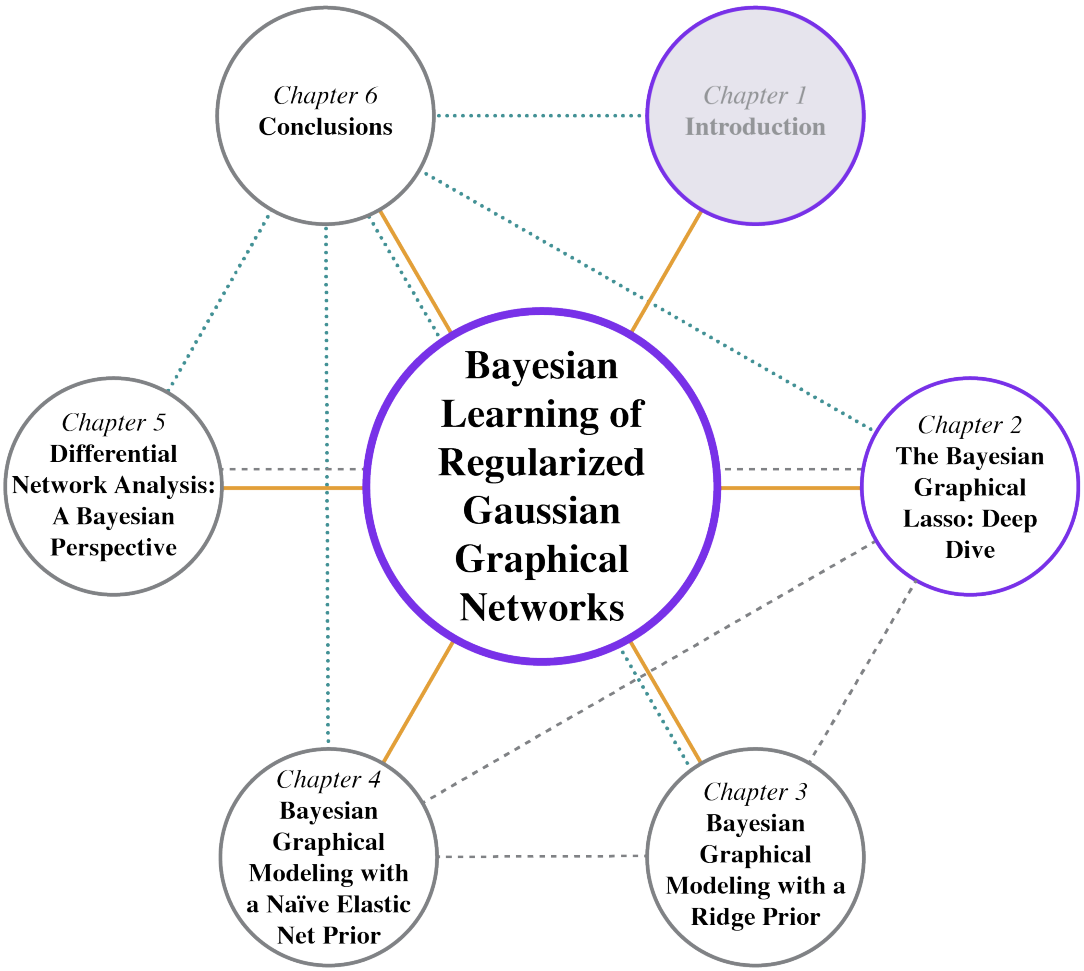
© University of Pretoria

© University of Pretoria

© University of Pretoria

Chapter 2

The Bayesian Graphical Lasso: Deep Dive



© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

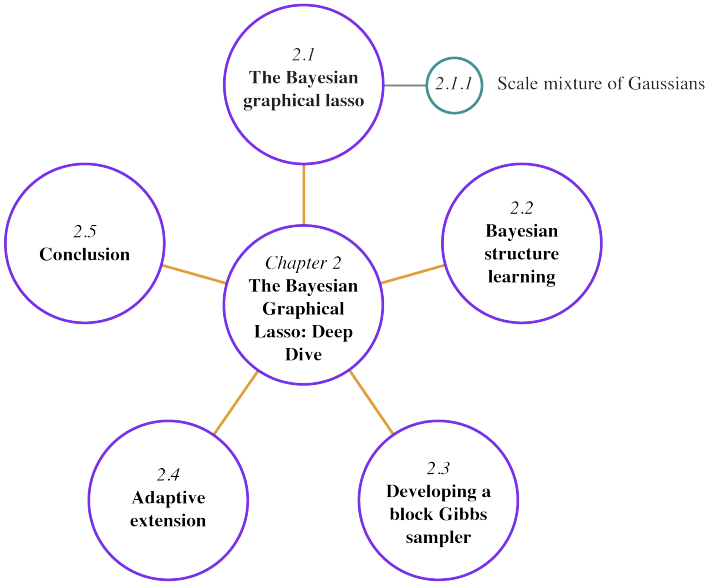


Fig. 2.1 *Chapter 2 visual outline.*

The technical endeavour of this thesis begins with an in-depth examination of the seminal work on the Bayesian graphical lasso presented in (Wang, 2012) - a critical foundation for the research presented in the subsequent chapters. This comprehensive review not only encapsulates the significant contribution of this methodology in the realm of Bayesian estimation of GGMs, but also underscores its foundational role as the bedrock upon which the methodologies developed in Chapters 3, 4, and 5 are constructed. That being said, the objective of the Bayesian graphical lasso is simple: provide a framework that enables efficient Bayesian inference for the graphical lasso model. Further numerical insights are provided in this chapter to enrich the contribution of (Wang, 2012). Figure 2.1 provides a visual outline of this chapter.

2.1 The Bayesian graphical lasso

The journey begins with the graphical lasso of (Friedman et al., 2008), a pioneering technique for estimating sparse inverse covariance matrices, thereby paving the way for significant advancements in the identification of GGMs. The objective of the graphical lasso is to maximise the penalised log-likelihood

$$\arg\max_{\boldsymbol{\Omega}\in\mathbb{M}^+}\left\{\log(\det\boldsymbol{\Omega})-\operatorname{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right)-\rho\|\boldsymbol{\Omega}\|_1\right\}. \tag{2.1}$$

Here, $\mathbf{S}=\mathbf{Y}'\mathbf{Y}$ denotes the $p\times p$ sample covariance matrix. The underlying assumption is that the observations $\mathbf{y}=(\mathbf{y}_1,\mathbf{y}_2,...,\mathbf{y}_n)$ are independent and identically distributed (iid) from a zero centered Gaussian distribution with covariance matrix $\boldsymbol{\Omega}^{-1}=\boldsymbol{\Sigma}$; where $\boldsymbol{\Omega}$ is contained in the space of positive definite matrices $\mathbb{M}^+$. Moreover, $\boldsymbol{\Omega}=(\omega_{ij})$ is the $p\times p$ precision matrix; $\rho\geq 0$ the shrinkage parameter and $\|\boldsymbol{\Omega}\|_1=\sum_{1\leq i,\,j\leq p}|\omega_{ij}|$ denotes the ℓ_1 norm of the precision matrix. Equation (2.1) is convex and the penalty parameter ρ is applied to all elements of $\boldsymbol{\Omega}$

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

and is typically chosen via cross-validation ((Friedman et al., 2008), (Meinshausen and Bühlmann, 2006)) or other model selection techniques such as Bayesian information criterion (BIC) in (Yuan and Lin, 2007).

The Bayesian connection to the graphical lasso is the maximum a posteriori (MAP) estimation of the following model

$$\begin{aligned} p\left(\mathbf{y}_i \mid \boldsymbol{\Omega}\right) &= \mathrm{N}_p\left(\mathbf{y}_i \mid \mathbf{0}, \boldsymbol{\Omega}^{-1}\right) \quad (i = 1, \ldots, n) \\ p\left(\boldsymbol{\Omega} \mid \lambda\right) &= C^{-1} \prod_{i<j} \left\{ \mathrm{DE}(\omega_{ij} \mid \lambda) \right\} \prod_{i=1}^{p} \left\{ \mathrm{EXP}(\omega_{ii} \mid \frac{\lambda}{2}) \right\} \mathbb{1}_{\boldsymbol{\Omega} \in \mathbb{M}^+}. \end{aligned} \tag{2.2}$$

In (2.2), $\mathrm{N}_p(\cdot)$ denotes a p-variate Gaussian probability density function (pdf) with form $p(y) = (\sigma\sqrt{2\pi})^{-1}\exp(-0.5[y/\sigma]^2)$. The prior for $\boldsymbol{\Omega}$ is given by the product of a double exponential (DE) with pdf $p(y) = \lambda/2\exp(-\lambda|y|)$ for the off-diagonal elements and an exponential (EXP) with pdf $p(y) = \lambda\exp(-\lambda y)\mathbb{1}_{y>0}$ for the diagonal. The indicator function $\mathbb{1}_{y>0}$ equals 1 if $y > 0$ and 0 otherwise. The normalising constant C does not involve λ - to see this, consider the following

$$C = \int_{\boldsymbol{\Omega} \in \mathbb{M}^+} \prod_{i<j} \left\{ \mathrm{DE}(\omega_{ij} \mid \lambda) \right\} \prod_{i=1}^{p} \left\{ \mathrm{EXP}(\omega_{ii} \mid \frac{\lambda}{2}) \right\} d\boldsymbol{\Omega}, \tag{2.3}$$

consider the change of variable $\tilde{\boldsymbol{\Omega}} = \lambda\boldsymbol{\Omega}$, it immediately follows that

$$\begin{aligned} C &= \int_{\tilde{\boldsymbol{\Omega}} \in \mathbb{M}^+} \prod_{i<j} \frac{\lambda}{2}\exp\left(-\lambda|g(\tilde{\omega}_{ij})|\right) \prod_{i=1}^{p} \frac{\lambda}{2}\exp\left(-\frac{\lambda}{2}g(\tilde{\omega}_{ii})\right) \frac{\partial(\tilde{\omega}_{ij})}{\partial(\omega_{ij})} d\tilde{\boldsymbol{\Omega}} \\ &= \int_{\tilde{\boldsymbol{\Omega}} \in \mathbb{M}^+} \prod_{i<j} \frac{1}{2}\exp\left(-|\tilde{\omega}_{ij}|\right) \prod_{i=1}^{p} \frac{1}{2}\exp\left(-\frac{1}{2}\tilde{\omega}_{ii}\right) d\tilde{\boldsymbol{\Omega}} \\ &= \int_{\tilde{\boldsymbol{\Omega}} \in \mathbb{M}^+} \prod_{i<j} \left\{ \mathrm{DE}(\tilde{\omega}_{ij} \mid 1) \right\} \prod_{i=1}^{p} \left\{ \mathrm{EXP}(\tilde{\omega}_{ii} \mid \frac{1}{2}) \right\} d\tilde{\boldsymbol{\Omega}}. \end{aligned} \tag{2.4}$$

The equality in (3.3) relies on the substitution $\tilde{\boldsymbol{\Omega}} = \lambda\boldsymbol{\Omega}$, where $g(\tilde{\omega}_{ij}) = \lambda\omega_{ij}$ for $i \leq j$ and notice that $\{\tilde{\boldsymbol{\Omega}} : \tilde{\boldsymbol{\Omega}} \in \mathbb{M}^+\} \equiv \{\boldsymbol{\Omega} : \boldsymbol{\Omega} \in \mathbb{M}^+\}$ for $\lambda > 0$. The posterior distribution associated with this prior is given by

$$\begin{aligned} p\left(\boldsymbol{\Omega} \mid \mathbf{Y}, \lambda\right) &\propto \prod_{i=1}^{n} \det\boldsymbol{\Omega}^{\frac{1}{2}} \exp\left\{ -\frac{1}{2}\mathbf{y}_i^T\boldsymbol{\Omega}\mathbf{y}_i \right\} \prod_{i<j} \frac{\lambda}{2}\exp\left(-\lambda|\omega_{ij}|\right) \prod_{i=1}^{p} \frac{\lambda}{2}\exp\left(-\frac{\lambda}{2}\omega_{ii}\right) \\ &\propto \det\boldsymbol{\Omega}^{\frac{n}{2}} \exp\left\{ -\frac{1}{2}\mathrm{trace}(\mathbf{S}\boldsymbol{\Omega}) - \sum_{i\neq j}\frac{\lambda}{2}|\omega_{ij}| - \sum_{i=1}^{p}\frac{\lambda}{2}\omega_{ii} \right\}. \end{aligned}$$

The log of the posterior distribution is given by

$$\begin{aligned} p_\ell\left(\boldsymbol{\Omega} \mid \mathbf{Y}, \lambda\right) &\propto \frac{n}{2}\log(\det\boldsymbol{\Omega}) - \frac{1}{2}\mathrm{trace}\left(\mathbf{S}\boldsymbol{\Omega}\right) - \sum_{i=1}^{p}\frac{\lambda}{2}\omega_{ii} - \sum_{i\neq j}\frac{\lambda}{2}|\omega_{ij}| \\ &\propto \log(\det\boldsymbol{\Omega}) - \mathrm{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right) - \sum_{i=1}^{p}\frac{\lambda}{n}\omega_{ii} - \sum_{i\neq j}\frac{\lambda}{n}|\omega_{ij}|. \end{aligned} \tag{2.5}$$

The maximum a posteriori (MAP) estimate is the value of $\boldsymbol{\Omega}$ which maximises (2.5), namely

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$\begin{aligned}\hat{\boldsymbol{\Omega}} &= \underset{\boldsymbol{\Omega}\in\mathbb{M}^+}{\arg\max}\left\{\log(\det\boldsymbol{\Omega}) - \operatorname{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right) - \sum_{1\leq i,\, j\leq p}|\omega_{ij}|\right\}\\ &= \underset{\boldsymbol{\Omega}\in\mathbb{M}^+}{\arg\max}\left\{\log(\det\boldsymbol{\Omega}) - \operatorname{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right) - \rho\|\boldsymbol{\Omega}\|_1\right\},\end{aligned}$$

where $\rho = \lambda/n$ and within the Bayesian context λ is treated as the shrinkage parameter. The block Gibbs sampler which is used to obtain $\hat{\boldsymbol{\Omega}}$ is discussed in the following sections.

2.1.1 Scale mixture of Gaussians

In the regression context, the Bayesian lasso (Park and Casella, 2008) exploits the double exponential distribution as a scale mixture of Gaussians (West, 1987) for the construction of an efficient block Gibbs sampler

$$\frac{\lambda}{2}\exp\left(-\lambda|x|\right) = \int_0^\infty \frac{1}{\sqrt{2\pi\tau}}\exp\left(-\frac{x^2}{2\tau}\right)\frac{\lambda^2}{2}\exp\left(-\frac{\lambda^2}{2}\tau\right)d\tau. \tag{2.6}$$

This representation assumes prior independence in the construction of their hierarchical representation of the full model and where τ is the scale parameter. The use of (2.6) for each off-diagonal element ω_{ij} in the graphical model domain seems fitting, unfortunately the positive definite constraint on $\boldsymbol{\Omega}$, invalidates the assumption of prior independence. In a similar fashion, the prior in (2.2) can be formulated as

$$p\left(\boldsymbol{\Omega}\,|\,\boldsymbol{\tau},\lambda\right) = C_\tau^{-1}\prod_{i<j}\left\{\frac{1}{\sqrt{2\pi\tau_{ij}}}\exp\left(-\frac{\omega_{ij}^2}{2\tau_{ij}}\right)\right\}\prod_{i=1}^{p}\left\{\frac{\lambda}{2}\exp\left(-\frac{\lambda}{2}\omega_{ii}\right)\right\}\mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^+}, \tag{2.7}$$

where $\boldsymbol{\tau} = \{\tau_{ij}\}_{i<j}$ is a vector of the scale parameters. The normalising constant, C_τ, has no closed-form solution. The mixing density for $\boldsymbol{\tau}$ is proportional to an exponential density with rate parameter $\lambda^2/2$

$$p(\boldsymbol{\tau}\,|\lambda) \propto C_\tau\prod_{i<j}\frac{\lambda^2}{2}\exp\left(-\frac{\lambda^2}{2}\tau_{ij}\right). \tag{2.8}$$

Simple substitution yields the intractable normalising constant in (2.7) and (2.8) simplify and the marginal distribution of the ω_{ij}'s adhere to (2.2). To see this, note that the marginals of ω_{ij} can be computed as follows

$$\begin{aligned}p\left(\omega\,|\,\lambda\right) = \int_0^\infty &\prod_{i<j}\left\{\frac{1}{\sqrt{2\pi\tau_{ij}}}\exp\left(-\frac{\omega_{ij}^2}{2\tau_{ij}}\right)\frac{\lambda^2}{2}\exp\left(-\frac{\lambda^2}{2}\tau_{ij}\right)\right\}\\ &\times\prod_{i=1}^{p}\left\{\frac{\lambda}{2}\exp\left(-\frac{\lambda}{2}\omega_{ii}\right)\right\}\mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^+}d(\tau_{ij})_{i\leq j}.\end{aligned} \tag{2.9}$$

Without loss of generality, consider the treatment of τ_{12} in the integrand of (2.9)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$
\begin{aligned}
f(\omega_{12}\,|\,\lambda) &= \int_0^\infty \frac{1}{\sqrt{2\pi\tau_{12}}}\exp\left\{-\frac{\omega_{12}^2}{2\tau_{12}}\right\}\frac{\lambda^2}{2}\exp\left\{-\frac{\lambda^2}{2}\tau_{12}\right\}d\tau_{12}\\
&= \frac{\lambda^2}{2}\int_0^\infty \frac{1}{\sqrt{2\pi\tau_{12}}}\exp\left\{-\frac{\lambda^2\tau_{12}}{2}-\frac{|\omega_{12}|^2}{2\tau_{12}}\right\}d\tau_{12}\\
&= \frac{\lambda^2}{2}\int_0^\infty \frac{1}{\sqrt{2\pi\tau_{12}}}\exp\left\{-\frac{\lambda^2}{2\tau_{12}}\left(\tau_{12}+\frac{|\omega_{12}|^2}{\lambda^2}\right)\right\}d\tau_{12}\\
&= \frac{\lambda^2}{2}\int_0^\infty \frac{1}{\sqrt{2\pi\tau_{12}}}\exp\left\{-\frac{\lambda^2}{2\tau_{12}}\left(\left[\tau_{12}-\frac{|\omega_{12}|}{\lambda}\right]^2+2\tau_{12}\frac{|\omega_{12}|}{\lambda}\right)\right\}d\tau_{12}\\
&= \frac{\lambda^2}{2}\exp\left\{-\lambda|\omega_{12}|\right\}\int_0^\infty \frac{1}{\sqrt{2\pi\tau_{12}}}\exp\left\{-\frac{\lambda^2}{2\tau_{12}}\left(\tau_{12}-\frac{|\omega_{12}|}{\lambda}\right)^2\right\}d\tau_{12}\\
&= \frac{\lambda^2}{2}\exp\left\{-\lambda|\omega_{12}|\right\}\int_0^\infty \frac{1}{\sqrt{2\pi\tau_{12}}}\exp\left\{-\frac{\lambda^2}{2\tau_{12}}\frac{[|\omega_{12}|/\lambda]^2}{[|\omega_{12}|/\lambda]^2}\left(\tau_{12}-\frac{|\omega_{12}|}{\lambda}\right)^2\right\}d\tau_{12}\\
&= \frac{\lambda^2}{2}\exp\left\{-\lambda|\omega_{12}|\right\}\int_0^\infty \frac{1}{\sqrt{2\pi\tau_{12}}}\exp\left\{-\frac{|\omega_{12}|^2}{2\tau_{12}[|\omega_{12}|/\lambda]^2}\left(\tau_{12}-\frac{|\omega_{12}|}{\lambda}\right)^2\right\}d\tau_{12}\\
&= \frac{\lambda^2}{2}\exp\left\{-\lambda|\omega_{12}|\right\}\frac{1}{|\omega_{12}|}\int_0^\infty \tau_{12}\sqrt{\frac{|\omega_{12}|^2}{2\pi\tau_{12}^3}}\exp\left\{-\frac{|\omega_{12}|^2}{2\tau_{12}[|\omega_{12}|/\lambda]^2}\left(\tau_{12}-\frac{|\omega_{12}|}{\lambda}\right)^2\right\}d\tau_{12},
\end{aligned}
$$

let $\gamma = |\omega_{12}|^2$ and $\mu = |\omega_{12}|/\lambda$

$$
f(\omega_{12}\,|\,\lambda) = \frac{\lambda^2}{2}\exp\left\{-\lambda|\omega_{12}|\right\}\frac{1}{\sqrt{\gamma}}\int_0^\infty \tau_{12}\sqrt{\frac{\gamma}{2\pi\tau_{12}^3}}\exp\left\{-\frac{\gamma}{2\tau_{12}\mu^2}(\tau_{12}-\mu)^2\right\}d\tau_{12},
$$

notice that the integral in the final step above represents the expected value (denoted by $\mathbb{E}(\cdot)$) of an inverse Gaussian distribution, see Chapter 2 of (Chhikara and Folks, 1989), it follows that

$$
\begin{aligned}
f(\omega_{12}\,|\,\lambda) &= \frac{\lambda^2}{2}\exp\left\{-\lambda|\omega_{12}|\right\}\frac{1}{|\omega_{12}|}\mathbb{E}(\tau_{12})\\
&= \frac{\lambda^2}{2}\exp\left\{-\lambda|\omega_{12}|\right\}\frac{1}{|\omega_{12}|}\mu\\
&= \frac{\lambda}{2}\exp\left\{-\lambda|\omega_{12}|\right\}.
\end{aligned}
$$

Notice that τ_{12} is no longer present in the integrand and can subsequently be treated as a constant in (2.9). Repeating this procedure for each $(\tau_{ij})_{i\leq j}$ implies that the marginal distribution of the $(\omega_{ij})_{i\leq j}$ follows (2.2).

Finally, the posterior distribution from which the Bayesian graphical lasso Gibbs sampler generates $\boldsymbol{\Omega}$ can be expressed as

$$
\begin{aligned}
p(\boldsymbol{\Omega},\boldsymbol{\tau}\,|\,\mathbf{Y},\lambda) \propto \det\boldsymbol{\Omega}^{\frac{n}{2}}\exp\left\{-\operatorname{trace}\left(\frac{1}{2}\mathbf{S}\boldsymbol{\Omega}\right)\right\}\prod_{i<j}\left\{\tau_{ij}^{-\frac{1}{2}}\exp\left(-\frac{\omega_{ij}^2}{2\tau_{ij}}\right)\exp\left(-\frac{\lambda^2}{2}\tau_{ij}\right)\right\}\\
\times\prod_{i=1}^{p}\left\{\exp\left(-\frac{\lambda}{2}\omega_{ii}\right)\right\}\mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^+}.
\end{aligned}
\tag{2.10}
$$

Figures 2.2a - 2.2c display the posterior median values for the first diagonal element (ω_{11}), the first off-diagonal element (ω_{12}) and the partial correlation value associated with the first off-diagonal element (ρ_{12}). Each sub-figure

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

considers $p \in \{5, 10, 25, 50, 75\}$ and the sample size n ranging from 5 to 500. The true $\mathbf{\Omega}$ is set to the identity matrix and results are generated from the posterior in (2.10), where $\lambda = 3$, using the MCMC sampling mechanism in Section 2.3. Notice how larger standard errors are evident across all figures (Figures 2.2a - 2.2c) for smaller sample sizes. The distribution of the median of the diagonal elements has an increasing trend as p increases, however, this effect is reduced given a large enough sample size. This aligns with behaviour of the marginal distributions of individual diagonal elements discussed in (Wang, 2012). This behaviour may adversely impact the accuracy of the precision estimates obtained using this estimator. Moreover, the distribution of the posterior medians of the off-diagonal and partial correlation elements are more tightly contained around zero, however, smaller sample size may negatively impact this. This implies that the prior favours sparsity of the off-diagonal elements as p increases. Figure 2.2d displays the 95% credible intervals for ρ_{12} for varying n and p. Smaller sample sizes have wider interval and every interval contains zero. These intervals can be used as a heuristic treatment for structure learning - see Section 3.6 of Chapter 3.

2.2 Bayesian structure learning

Recall from Chapter 1 that, out of the box, the Bayesian graphical lasso is incapable of structure learning since it is constructed of absolute continuous priors. This formulation places zero probability mass on the event $\{\omega_{ij} = 0\}$ and so the posterior assigns zero probability to these events. In this case, structure learning requires a heuristic treatment. An adaption of the thresholding rule for the horseshoe estimator in (Carvalho et al., 2010) is used here. In particular, the Bayes posterior mean estimator of each partial correlation element ρ_{ij} is

$$\hat{\rho}_{ij} = \hat{\theta}_{ij}\mathbb{E}_g(\rho_{ij} \mid \mathbf{Y}),$$

where $\hat{\theta}_{ij}$ is the probability $\omega_{ij} \neq 0$, g is the distribution of the nonzero ω_{ij} and $\mathbb{E}$ denoting the expected value operator. Here $\hat{\theta}_{ij}$ serves a dual role. First, as the posterior edge inclusion probability, determining which connections are included in the graphical model. Second, as the amount of shrinkage applied to ρ_{ij}, controlling the reduction of the relationship strength between variables. Focusing on the latter, the posterior mean estimator of ρ_{ij} under the Bayesian graphical lasso prior is given by

$$\tilde{\rho}_{ij} = \tilde{\theta}_{ij}\mathbb{E}_g(\rho_{ij} \mid \mathbf{Y}),$$

where, $\tilde{\theta}_{ij}$ describes the shrinkage applied by the Bayesian graphical lasso prior on $\mathbb{E}_g(\rho_{ij} \mid \mathbf{Y})$. Finally, analogous to the decision rule for determining whether $\omega_{ij} \neq 0$ under a symmetric 0-1 loss function for the horseshoe prior in (Carvalho et al., 2010), the two shrinkage parameters $\hat{\theta}_{ij}$ and $\tilde{\theta}_{ij}$ can be linked so that (Wang, 2012) may claim $\omega_{ij} \neq 0$ if

$$\tilde{\theta}_{ij} = \frac{\tilde{\rho}_{ij}}{\mathbb{E}_g(\rho_{ij} \mid \mathbf{Y})} > 0.5. \tag{2.11}$$

Here, $\tilde{\rho}_{ij}$ is the posterior sample mean estimate of the partial correlation associated with the Bayesian graphical lasso prior and g is the standard conjugate Wishart prior $W(3, \mathbf{I}_p)$. The degrees of freedom parameter is selected based on the simulated experiments and gene expression application in (Jones et al., 2005). The identity matrix $\mathbf{I}_p$ is used as the positive definite scale matrix implying that the authors assume that the off-diagonal elements of $\mathbf{\Omega}$ are centered around zero. It is well known that the conjugate Wishart prior sets the gold standard for $\mathbf{\Omega}$ modelling. That being said, at the time of writing the marginal prior for partial correlations using a Wishart prior for $\mathbf{\Omega}$ has not been thoroughly investigated. While the marginal prior for ρ_{ij} cannot be expressed in a closed form, the

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

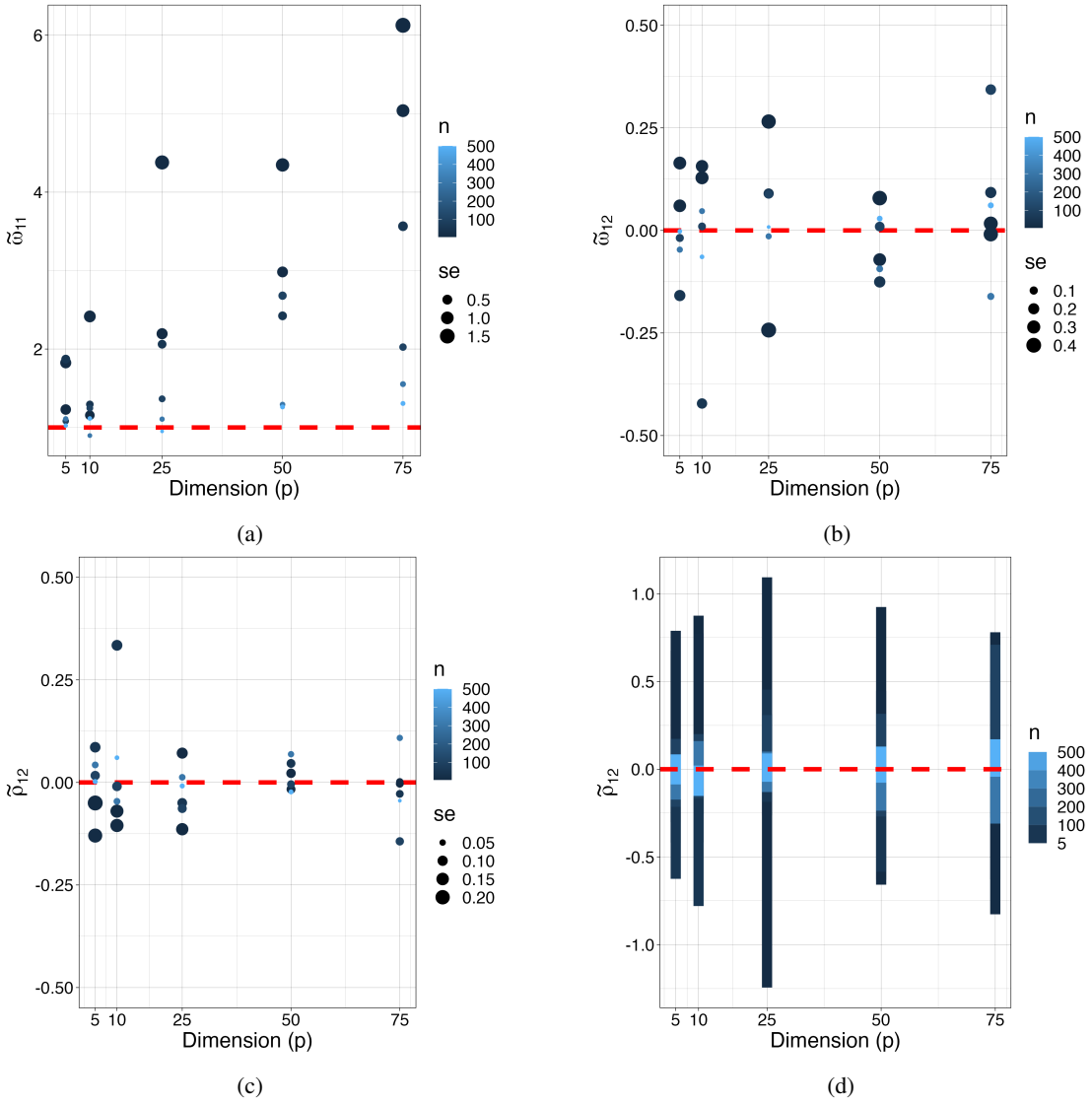


Fig. 2.2 *Distribution of the posterior median for a diagonal (a), an off-diagonal (b), partial correlation (c) entry of* $\mathbf{\Omega}$ *for* $p \in \{5, 10, 25, 50, 75\}$ *and the sample size n ranging from 5 to 500.* The size of dots are proportional to the standard error (se) of the elements and the red line indicates the true value. Credible intervals for the partial correlation entry are provided in (d). The red horizontal dashed line indicates the true value and the true precision matrix is set to the identity matrix.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

moments of the Wishart can be conveniently accessed (Drton et al., 2008). To this end, the implied prior can be used by utilising established properties of the Wishart distribution and numerically generating samples from it. In other words, given the aforementioned Wishart prior, the corresponding posterior of $\mathbf{\Omega}$ is given by a Wishart $W(3+n, (n\mathbf{S}+\mathbf{I}_p)^{-1})$. To see this, suppose the hierarchical form of $\mathbf{\Omega}$ is given by

$$\begin{aligned} p(\mathbf{y}_i \,|\, \mathbf{\Omega}) &= \mathrm{N}_p\left(\mathbf{y}_i \,|\, \mathbf{0}, \mathbf{\Omega}^{-1}\right) \quad (i = 1, \ldots, n) \\ p(\mathbf{\Omega} \,|\, \delta, \mathbf{D}) &= \frac{1}{2^{\delta p/2}} \frac{1}{|\mathbf{D}|^{\delta/2}\Gamma_p(\delta/2)} |\mathbf{\Omega}|^{\frac{\delta-p-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{trace}(\mathbf{D}^{-1}\mathbf{\Omega})\right\}, \end{aligned} \tag{2.12}$$

where Γ_p is the multivariate gamma function and $\mathbf{D}$ is a $p \times p$ positive definite symmetric scale matrix, see Chapter 3 of (Muirhead, 1982). The posterior is given by

$$p(\mathbf{\Omega} \,|\, \mathbf{Y}) \propto p(\mathbf{\Omega}) p(\mathbf{Y} \,|\, \mathbf{\Omega}),$$

or more specifically

$$\begin{aligned} p(\mathbf{\Omega} \,|\, \mathbf{Y}, \delta, \mathbf{D}) &\propto \prod_{i=1}^{n} |\mathbf{\Omega}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{y}_i^T \mathbf{\Omega} \mathbf{y}_i\right\} \times \\ &\quad \frac{1}{2^{\delta p/2}} \frac{1}{|\mathbf{D}|^{\delta/2}\Gamma_p(\delta/2)} |\mathbf{\Omega}|^{\frac{\delta-p-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{trace}(\mathbf{D}^{-1}\mathbf{\Omega})\right\} \text{ (see (2.12))} \\ &\propto \frac{1}{2^{(\delta+n)p/2}} \frac{1}{|\mathbf{D}|^{(\delta+n)/2}\Gamma_p((\delta+n)/2)} |\mathbf{\Omega}|^{\frac{\delta+n-p-1}{2}} \exp\left\{-\frac{1}{2}\left(\mathrm{trace}(\mathbf{\Omega}[n\mathbf{S}+\mathbf{D}^{-1}])\right)\right\}. \end{aligned}$$

Clearly the posterior is again a Wishart distribution $W(\delta+n, (n\mathbf{S}+\mathbf{D}^{-1})^{-1})$. Moreover, $\mathbf{S}$ is the sum of squares matrix $\mathbf{Y}^T\mathbf{Y}$ as seen earlier and the posterior mode and mean have a closed form, respectively

$$\arg\max_{\mathbf{\Omega}} p(\mathbf{\Omega} \,|\, \mathbf{Y}, \delta, \mathbf{D}) = (\delta+n-p-1)(n\mathbf{S}+\mathbf{D}^{-1})^{-1}, \tag{2.13}$$

$$\mathbb{E}(\mathbf{\Omega} \,|\, \mathbf{Y}, \delta, \mathbf{D}) = (\delta+n)(n\mathbf{S}+\mathbf{D}^{-1})^{-1}. \tag{2.14}$$

Finally, given that $\mathbf{\Omega}$ is directly obtainable from (2.14), the partial correlation matrix $\mathbf{P}$ can be computed as

$$\mathbb{E}(\mathbf{P} \,|\, \mathbf{Y}, \delta, \mathbf{D}) = -\left([\mathrm{diag}(\boldsymbol{\omega})]^{-1}\mathbf{\Omega}[\mathrm{diag}(\boldsymbol{\omega})]^{-1} - \mathbf{D}^{-1}\right), \tag{2.15}$$

where $\mathrm{diag}(\boldsymbol{\omega})$ represents a $p \times p$ diagonal matrix with $\sqrt{\omega_{ii}}$ for $i = 1, \cdots, p$ as the diagonal elements. For this dissertation, the non-informative prior $\mathbf{D} = \mathbf{I}_p$ is considered.

2.3 Developing a block Gibbs sampler

The block Gibbs sampler described here enables posterior computation of $\mathbf{\Omega}$ from the posterior distribution in (2.10). The Gibbs sampler sequentially updates $\mathbf{\Omega}$ one column at time. Without loss of generality, consider updating the last column and row of $\mathbf{\Omega}$ and let $\mathbf{\Upsilon}$ be a symmetric $p \times p$ matrix with a zero main diagonal and $\boldsymbol{\tau}$ in the upper and lower off-diagonal entries. Here, $\boldsymbol{\tau}$ represents a vector of the scale parameter τ_{ij} values from the scale mixture of Gaussian representation (see Subsection 2.1.1). Partition the matrices $\mathbf{\Omega}$, $\mathbf{S}$ and $\mathbf{\Upsilon}$ as follows

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$\underset{p\times p}{\boldsymbol{\Omega}} = \begin{pmatrix} \underset{(p-1)\times(p-1)}{\boldsymbol{\Omega}_{11}} & \underset{(p-1)\times 1}{\boldsymbol{\omega}_{12}} \\ \underset{1\times(p-1)}{\boldsymbol{\omega}_{21}} & \underset{1\times 1}{\omega_{22}} \end{pmatrix}, \quad \underset{p\times p}{\boldsymbol{S}} = \begin{pmatrix} \underset{(p-1)\times(p-1)}{\boldsymbol{S}_{11}} & \underset{(p-1)\times 1}{\mathbf{s}_{12}} \\ \underset{1\times(p-1)}{\mathbf{s}_{21}} & \underset{1\times 1}{s_{22}} \end{pmatrix}, \quad \underset{p\times p}{\boldsymbol{\Upsilon}} = \begin{pmatrix} \underset{(p-1)\times(p-1)}{\boldsymbol{\Upsilon}_{11}} & \underset{(p-1)\times 1}{\boldsymbol{\tau}_{12}} \\ \underset{1\times(p-1)}{\boldsymbol{\tau}_{21}} & \underset{1\times 1}{0} \end{pmatrix}. \tag{2.16}$$

Recall that from the properties of the determinant of block matrices (Silvester, 2000), $\det\boldsymbol{\Omega}^{\frac{n}{2}}$ can be represented as

$$\det\boldsymbol{\Omega}^{\frac{n}{2}} = (\omega_{22} - \boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}} \det\boldsymbol{\Omega}_{\mathbf{11}}^{\frac{n}{2}} \propto (\omega_{22} - \boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}},$$

since we are only interested in the last column and row. Similarly,

$$\begin{aligned} \operatorname{trace}\left(\frac{1}{2}\mathbf{S}\boldsymbol{\Omega}\right) &= \operatorname{trace}\left\{\frac{1}{2}\begin{pmatrix} \boldsymbol{S}_{11}\boldsymbol{\Omega}_{11} + \mathbf{s}_{12}\boldsymbol{\omega}_{21} & \boldsymbol{S}_{11}\boldsymbol{\omega}_{12} + \mathbf{s}_{12}\omega_{22} \\ \mathbf{s}_{21}\boldsymbol{\Omega}_{11} + s_{22}\boldsymbol{\omega}_{21} & \mathbf{s}_{21}\boldsymbol{\omega}_{12} + s_{22}\omega_{22} \end{pmatrix}\right\} \\ &\propto -\frac{1}{2}(2\mathbf{s}_{21}\boldsymbol{\omega}_{12} + s_{22}\omega_{22}). \end{aligned}$$

The last two approximations stem from the fact that we are only interested in updating the last column. Notice that the conditional distribution of the last column in (2.10) can be presented as

$$\begin{aligned} p(\boldsymbol{\omega}_{\mathbf{12}}, \omega_{22} \,|\, \boldsymbol{\Omega}_{11}, \boldsymbol{\Upsilon}, \mathbf{S}, \lambda) &\propto (\omega_{22} - \boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}} \exp\left(\frac{1}{2}(2\mathbf{s}_{21}\boldsymbol{\omega}_{12} + s_{22}\omega_{22})\right) \\ &\quad \times \prod_{i<j}\left\{\exp\left(-\frac{\omega_{ij}^2}{2\tau_{ij}}\right)\right\}\prod_{i=1}^{p}\left\{\exp\left(-\omega_{ii}\frac{\lambda}{2}\right)\right\}\mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^+} \\ &\propto (\omega_{22} - \boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}} \exp\left(\frac{1}{2}(2\mathbf{s}_{21}\boldsymbol{\omega}_{12} + s_{22}\omega_{22})\right) \\ &\quad \times \exp\left(-\frac{1}{2}\left\{\sum_{i=1}^{p-1}\frac{\omega_{ip}^2}{\tau_{ip}} + \omega_{22}\lambda\right\}\right) \text{ (we only focus on } \boldsymbol{\omega}_{\mathbf{12}} \text{ and } \omega_{22}) \\ &= (\omega_{22} - \boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}} \\ &\quad \times \exp\left(-\frac{1}{2}\left\{\boldsymbol{\omega}_{21}\mathbf{D}_{\boldsymbol{\tau}}^{-1}\boldsymbol{\omega}_{12} + 2\mathbf{s}_{21}\boldsymbol{\omega}_{12} + \omega_{22}(\lambda + s_{22})\right\}\right), \end{aligned}$$

where $\mathbf{D}_{\boldsymbol{\tau}}^{-1} = \operatorname{diag}(\boldsymbol{\tau}_{12})$. Next, consider the following change of variables

$$\begin{aligned} \boldsymbol{\beta} &= \boldsymbol{\omega}_{12} \\ \gamma &= \omega_{22} - \boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}, \end{aligned}$$

with the constant Jacobian not involving $\boldsymbol{\beta}$ nor γ. To see this, note that the transformation can also be viewed as follows

$$\begin{aligned} y_i &= \beta_{ip} = \omega_{ip} \text{ for } 1 \le i \le (p-1) \\ y_p &= \gamma = \omega_{22} - \boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12} \\ &= \omega_{22} - \sum_{i=1}^{p-1}\sum_{j=1}^{p-1}\omega_{pi}\omega_{ij}^*\omega_{jp}, \ \boldsymbol{\Omega}_{11}^{-1} = (\omega_{ij}^*). \end{aligned}$$

It follows that

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$\omega_{ip} = g_i(y_i, y_p) = y_i \text{ for } 1 \leq i \leq (p-1)$$
$$\omega_{22} = g_p(y_i, y_p) = y_p + \sum_{i=1}^{p-1}\sum_{j=1}^{p-1} y_i \omega^*_{ij} y_j \text{ for } 1 \leq i \leq (p-1),$$

and the Jacobian is given by

$$J = \det\begin{pmatrix} \frac{\partial g_1(y_1,y_p)}{\partial y_1} & \cdots & \frac{\partial g_1(y_1,y_p)}{\partial y_{(p-1)}} & \frac{\partial g_1(y_1,y_p)}{\partial y_p} \\ \vdots & \ddots & \cdots & \vdots \\ \frac{\partial g_{(p-1)}(y_{(p-1)},y_p)}{\partial y_1} & \cdots & \frac{\partial g_{(p-1)}(y_{(p-1)},y_p)}{\partial y_{(p-1)}} & \frac{\partial g_{(p-1)}(y_{(p-1)},y_p)}{\partial y_p} \\ \frac{\partial g_p(y_i,y_p)_{1\leq i\leq (p-1)}}{\partial y_1} & \cdots & \frac{\partial g_p(y_i,y_p)_{1\leq i\leq (p-1)}}{\partial y_{(p-1)}} & \frac{\partial g_p(y_i,y_p)_{1\leq i\leq (p-1)}}{\partial y_p} \end{pmatrix}$$

$$= \det\begin{pmatrix} 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & 1 & 0 \\ 2y_1\omega^*_{11} & \cdots & 2y_{(p-1)}\omega^*_{(p-1)(p-1)} & 1 \end{pmatrix}.$$

For the simple case where $p = 3$, the Jacobian is given by

$$J = \det\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2y_1\omega^*_{11} & 2y_2\omega^*_{22} & 1 \end{pmatrix}$$
$$= \det\begin{pmatrix} 1 & 0 \\ 2y_2\omega^*_{22} & 1 \end{pmatrix}$$
$$= 1.$$

Clearly the Jacobian is a constant not involving $\boldsymbol{\beta}$ nor γ. Finally, this transformation yields the following conditional distribution

$$p(\gamma, \boldsymbol{\beta} | \boldsymbol{\Omega}_{11}, \boldsymbol{\Upsilon}, \boldsymbol{Y}, \lambda) \propto (\gamma)^{\frac{n}{2}} \exp\left(-\frac{1}{2}\left\{\boldsymbol{\beta}^T \mathbf{D}_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta} + 2\mathbf{s}_{21}\omega_{12} + (\gamma + \boldsymbol{\beta}^T\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\beta})(\lambda + s_{22})\right\}\right)$$
$$= (\gamma)^{\frac{n}{2}} \exp\left(-\frac{s_{22}+\lambda}{2}\gamma\right) \exp\left(-\frac{1}{2}[\boldsymbol{\beta}^T\{\mathbf{D}_{\boldsymbol{\tau}}^{-1} + (s_{22}+\lambda)\boldsymbol{\Omega}_{11}^{-1}\}\boldsymbol{\beta} + 2\mathbf{s}_{21}\boldsymbol{\beta}]\right).$$

It follows that

$$(\gamma, \boldsymbol{\beta} | \boldsymbol{\Omega}_{11}, \boldsymbol{\Upsilon}, \boldsymbol{Y}, \lambda) \sim \text{GA}\left(\frac{n}{2}+1, \frac{s_{22}+\lambda}{2}\right)\text{N}_{p-1}(-\boldsymbol{C}\boldsymbol{s}_{21}, \boldsymbol{C}),$$

where $\boldsymbol{C} = \{(s_{22}+\lambda)\boldsymbol{\Omega}_{11}^{-1} + \mathbf{D}_{\boldsymbol{\tau}}^{-1}\}^{-1}$, GA$(\cdot)$ refers to a gamma pdf and N$_{p-1}(\cdot)$ to a $(p-1)$-variate Gaussian pdf.

Recall that a positive definite matrix requires all leading principal minors to be positive. Here, the leading principal minors of a matrix are the determinants of the leading principal matrices. The leading principal matrices of a $p \times p$ matrix $\mathbf{M}$ (the choice of 'M' is purely for illustration) are given by the set $\{D_i : 1 \leq i \leq p\}$ where D_i is obtained by removing the last $p-i$ rows and columns from $\mathbf{M}$. That being said, let $\boldsymbol{\Omega}^{(k)}$ be the kth positive definite sample from the block Gibbs sampler defined above. In other words, all the leading principal minors of $\boldsymbol{\Omega}^{(k)}$ are positive. Next, in the $(k+1)^{th}$ iteration, the final column and row of $\boldsymbol{\Omega}^{(k)}$ is updated to produce $\boldsymbol{\Omega}^{(k+1)}$. The leading principal matrices of $\boldsymbol{\Omega}^{(k)}$ and $\boldsymbol{\Omega}^{(k+1)}$ are identical up to $D_{(p-1)}$, in other words the only leading

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

principal matrices which differ in this iteration are the full $\mathbf{\Omega}^{(k)}$ and $\mathbf{\Omega}^{(k+1)}$. It is also clear that $\det(\mathbf{\Omega}^{(k+1)}) = \gamma\det(\mathbf{\Omega}_{11}^{(k)})$, where $\mathbf{\Omega}_{11}^{(k)}$ is the $(p-1)^{th}$ leading principal matrix of $\mathbf{\Omega}^{(k)}$, which we know is positive. It is clear that since $\gamma > 0$, it is true that $\det(\mathbf{\Omega}^{(k+1)}) > 0$ and as a result $\mathbf{\Omega}^{(k+1)}$ is positive definite. To see this more clearly with an illustrative example, refer to Subsection A.1.1 in the appendix.

Updating $\boldsymbol{\tau}$ is achieved by noticing that the conditional posterior distribution of $1/\tau_{ij}$ is independently inverse Gaussian $\text{IG}(\mu^*, \lambda^*)$ where $\mu^* = \sqrt{\lambda^2/\omega_{ij}^2}$ and $\lambda^* = \lambda^2$. To see this more clearly, note that the hierarchical form in (2.7) and (2.8) can be used to derive the conditional posterior distributions of the $\psi_{ij} = 1/\tau_{ij}$. Therefore, updating each τ_{ij} in $\boldsymbol{\tau}$ from (2.10) is equivalent to sampling from

$$\begin{aligned} p\left(\psi_{ij} \,|\, \mathbf{\Omega}, \mathbf{Y}, \lambda\right) &= \psi_{ij}^{\frac{1}{2}} \exp\left(-\frac{\omega_{ij}^2\psi_{ij}}{2}\right)\exp\left(-\frac{\lambda^2}{2\psi_{ij}}\right) \\ &= \psi_{ij}^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left[\omega_{ij}^2\psi_{ij} + \frac{\lambda^2}{\psi_{ij}}\right]\right). \end{aligned} \tag{2.17}$$

The form in (2.17) happens to be proportional to the probability density function of the reciprocal of an inverse Gaussian random variable. More specifically

$$\begin{aligned} p\left(\psi_{ij} \,|\, \mathbf{\Omega}, \mathbf{Y}, \lambda\right) &\propto \psi_{ij}^{-\frac{3}{2}} \exp\left(-\frac{1}{2}\left[\omega_{ij}^2\psi_{ij} + \frac{\lambda^2}{\psi_{ij}}\right]\right) \\ &= \psi_{ij}^{-\frac{3}{2}} \exp\left(-\frac{\lambda^2}{2}\left\{\frac{\omega_{ij}^2\psi_{ij}}{\lambda^2} + \frac{1}{\psi_{ij}}\right\}\right) \\ &= \psi_{ij}^{-\frac{3}{2}} \exp\left(-\lambda^2\left\{\frac{\psi_{ij}^2 + \left(\sqrt{\lambda^2/\omega_{ij}^2}\right)^2}{2\left(\sqrt{\lambda^2/\omega_{ij}^2}\right)^2\psi_{ij}}\right\}\right) \\ &\propto \psi_{ij}^{-\frac{3}{2}} \exp\left(-\frac{\lambda^2\left(\psi_{ij} - (\lambda^2/\omega_{ij}^2)\right)^2}{2\left(\sqrt{\lambda^2/\omega_{ij}^2}\right)^2\psi_{ij}}\right), \end{aligned} \tag{2.18}$$

here, the form given in (2.18) is comparable with the popular parameterisation of the inverse Gaussian pdf proposed in (Chhikara and Folks, 1989), namely

$$p(x; \mu^*, \lambda^*) = \sqrt{\frac{\lambda^*}{2\pi}}x^{-\frac{3}{2}}\exp\left(-\frac{\lambda^*(x-\mu^*)^2}{2(\mu^*)^2x}\right) \quad x > 0.$$

Finally, the conditional posterior distributions for the ψ_{ij} is given by the inverse Gaussian pdf where $\mu* = \sqrt{\lambda^2/\omega_{ij}^2}$ and $\lambda^* = \lambda^2$. Conclusively, the block Gibbs sampler is summarised in Algorithm 1.

2.4 Adaptive extension

Revisiting the double exponential prior in (2.2), it is important to acknowledge its known restrictions. It has a tendency to over-shrink larger coefficients while not shrinking smaller ones enough. These tendencies have been extensively explored in the fields of regression and graphical modelling, as evidenced by works such as (Brown and Griffin, 2010; Carvalho et al., 2010), sparking the creation of alternate priors. The hierarchical form in (2.7), (2.8) and the block Gibbs sampler in Algorithm 1 may be tailored to allow for adaptive shrinkage on the off-diagonal

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Algorithm 1 Bayesian graphical lasso: Block Gibbs sampler

Require: Initialise $\boldsymbol{\Omega} \in \mathbb{M}^+$.
  for $i = 1, \ldots, p$ **do**
    1) Partition $\boldsymbol{\Omega}$, $\mathbf{S}$ and $\boldsymbol{\Upsilon}$ as in (4.7).
    2) Sample $\gamma \sim \text{GA}(n/2+1, (s_{22}+\lambda)/2)$ and $\boldsymbol{\beta} \sim \text{N}_{\text{p}-1}(-\boldsymbol{C}\boldsymbol{s}_{21}, \boldsymbol{C})$.
    3) Update $\boldsymbol{\omega}_{12} = \boldsymbol{\beta}$, $\boldsymbol{\omega}_{21} = \boldsymbol{\beta}^\top$ and $\omega_{22} = \gamma + \boldsymbol{\beta}^\top \boldsymbol{\Omega}_{\mathbf{11}}{}^{-1}\boldsymbol{\beta}$.
  end for
  for $i \neq j$ **do**
    Sample $\psi_{ij} \sim \text{IG}(\mu^*, \lambda^*)$ where $\lambda^* = \lambda^2$ and $\mu^* = \sqrt{(\lambda^2/\omega_{ij}^2)}$ and update $\tau_{ij} = 1/\psi_{ij}$.
  end for

elements ω_{ij}. To this end, the Bayesian adaptive graphical lasso is introduced which draws inspiration from the frequentist adaptive graphical lasso. The penalised log likelihood of the latter is given by

$$\log(\det\boldsymbol{\Omega}) - \text{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right) - \lambda \sum_{i=1}^{p}\sum_{j=1}^{p} w_{ij}|\omega_{ij}|, \tag{2.19}$$

where w_{ij} function as the adaptive weights and recommend $w_{ij} = 1/|\omega_{ij}|^\alpha$ for $\alpha > 0$ (Fan et al., 2009). The Bayesian equivalent of (2.19) is given by the set of priors

$$\begin{aligned} p(\boldsymbol{\Omega} \,|\, \{\lambda_{ij}\}_{i\leq j}) &= C^{-1}_{\{\lambda_{ij}\}_{i\leq j}} \prod_{i<j}\left\{\text{DE}(\omega_{ij}|\lambda_{ij})\right\}\prod_{i=1}^{p}\left\{\text{EXP}(\omega_{ii}|\frac{\lambda_{ii}}{2})\right\}\mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^+} \\ p(\{\lambda_{ij}\}_{i<j} \,|\, \{\lambda_{ii}\}_{i=1}^{p}) &\propto C_{\{\lambda_{ij}\}_{i\leq j}} \prod_{i<j}\text{GA}(r,s). \end{aligned} \tag{2.20}$$

The set of priors in (2.20) make use of different shrinkage parameters λ_{ij} on each off-diagonal element ω_{ij}. Secondly, $\{\lambda_{ii}\}_{i=1}^{p}$ for the diagonal elements are treated as hyperparameters. Finally, $C_{\{\lambda_{ij}\}_{i\leq j}}$ is the intractable normalisation constant dependent on λ_{ij}, however, these constants will simplify in the formulation of the posterior distribution - see Subsection 2.1.1. The posterior of the $\{\lambda_{ij}\}_{i<j}$ conditioned on $\boldsymbol{\Omega}$ is given by

$$\begin{aligned} p(\lambda_{ij} \,|\, \boldsymbol{\Omega}) &= \frac{\lambda_{ij}}{2}\exp(-\lambda_{ij}|\omega_{ij}|)\frac{s^r}{\Gamma(r)}\lambda_{ij}^{r-1}\exp(-s\lambda_{ij}) \\ &\propto \frac{s^{r+1}}{\Gamma(r+1)}\lambda_{ij}^{r}\exp\Big\{-\lambda_{ij}(|\omega_{ij}|+s)\Big\}, \end{aligned} \tag{2.21}$$

which is the pdf of a gamma random variable with shape parameter $(r+1)$ and scale parameter $(|\omega_{ij}|+s)$. Clearly the conditional posterior is given by $\text{GA}(r+1, |\omega_{ij}|+s)$. Notice that the mean of λ_{ij} is $(r+1)/(|\omega_{ij}|+s)$. In other words, the amount of shrinkage applied by λ_{ij} depends on the magnitude of ω_{ij}, however, the shape and scale parameters (r, s) should be small enough to ensure the adaptiveness is mainly controlled by the value of ω_{ij}. The updated block Gibbs sampler is presented in Algorithm 2.

2.5 Conclusion

This chapter establishes the essential groundwork for the Bayesian graphical lasso and its adaptive variant, including their corresponding block Gibbs samplers. These foundational components underpin the subsequent chapters' investigations. When striving for shrinkage toward zero, the Bayesian graphical lasso emerges as a strong

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Algorithm 2 Bayesian adaptive graphical lasso: Block Gibbs sampler

Require: Initialise $\boldsymbol{\Omega} \in \mathbb{M}^+$, $\{\lambda_{ii}\}_{i=1}^{p}$, r and s.
  for $i = 1, \ldots, p$ **do**
    1) Partition $\boldsymbol{\Omega}$, $\mathbf{S}$ and $\boldsymbol{\Upsilon}$ as in (4.7).
    2) Sample $\gamma \sim \text{GA}(n/2+1, (s_{22}+\lambda_{ii})/2)$ and $\boldsymbol{\beta} \sim \text{N}_{\text{p}-1}(-\boldsymbol{C}\boldsymbol{s}_{21}, \boldsymbol{C})$.
    3) Update $\boldsymbol{\omega}_{12} = \boldsymbol{\beta}$, $\boldsymbol{\omega}_{21} = \boldsymbol{\beta}^\top$ and $\omega_{22} = \gamma + \boldsymbol{\beta}^\top \boldsymbol{\Omega_{11}}^{-1} \boldsymbol{\beta}$.
  end for
  for $i \neq j$ **do**
    1) Sample $\lambda_{ij} \sim \text{GA}(r+1, |\omega_{ij}| + s)$.
    2) Sample $\psi_{ij} \sim \text{IG}(\mu^*, \lambda^*)$ where $\lambda^* = \lambda_{ij}^2$ and $\mu^* = \sqrt{(\lambda_{ij}^2/\omega_{ij}^2)}$ and update $\tau_{ij} = 1/\psi_{ij}$.
  end for

contender. Its prior progressively favours partial correlation values near 0 with increasing dimensionality (Wang, 2012). However, in cases where extreme sparsity in the true precision matrix is not imperative, there is merit in favouring regularisation techniques that proportionally shrink estimated elements (Wieringen and Peeters, 2016). This precise focus defines the upcoming chapter's scope.

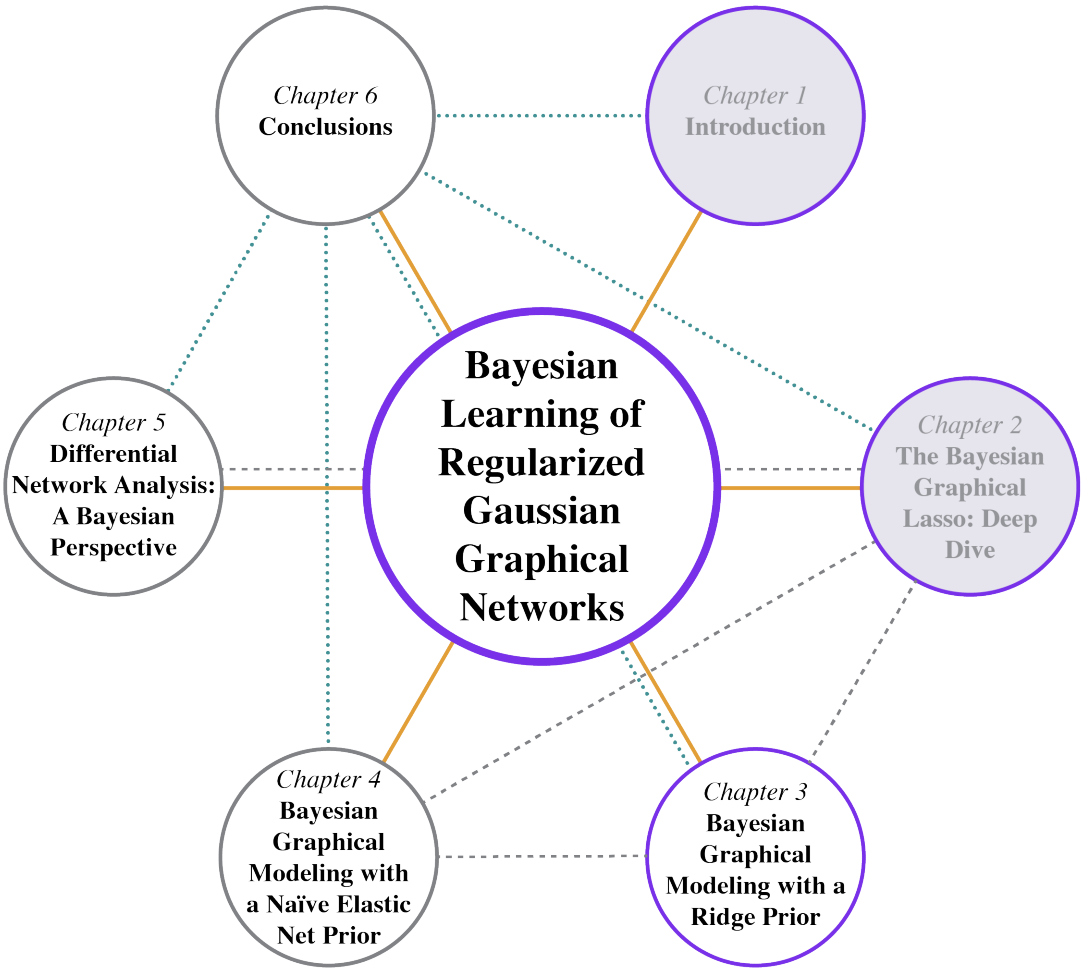
© University of Pretoria

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chapter 3

Bayesian Graphical Modeling with a Ridge Prior



© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

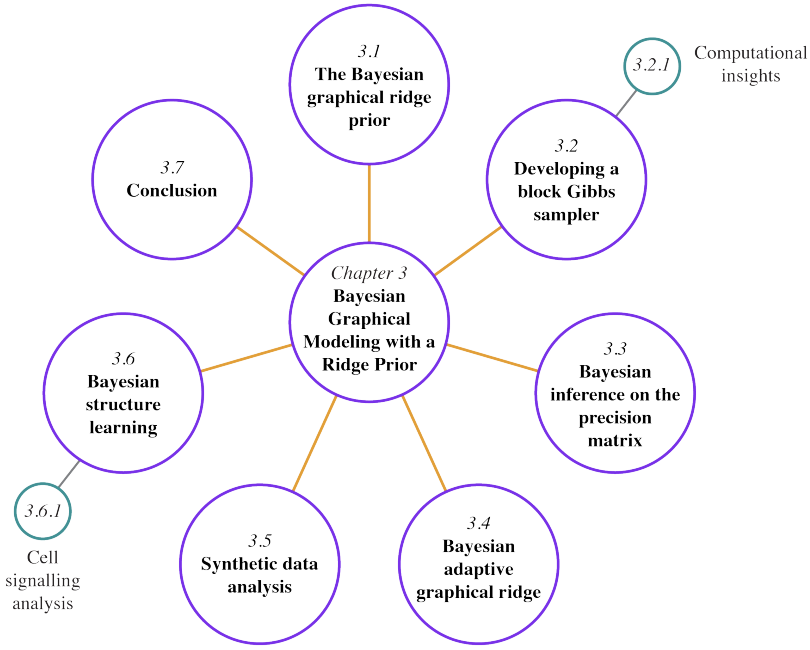


Fig. 3.1 *Chapter 3 visual outline.*

Juxtaposed to situations where sparsity proves advantageous, there may exist scenarios which demand a more profound pursuit of accurate representations of the conditional (in)dependencies encoded within precision matrices. Moreover, the true (graphical) model need not exhibit extensive sparsity. In such instances, prioritisation on regularisation methods which shrink the estimated elements of the precision matrix proportionally, in conjunction with a heuristic structure learning treatment seems reasonable Wieringen and Peeters (2016). This chapter builds upon the foundations of the Bayesian graphical lasso in Chapter 2 and introduces a novel framework that enables efficient Bayesian inference for the graphical models under a ridge type penalisation. Figure 3.1 provides a visual outline of this chapter.

3.1 The Bayesian graphical ridge prior

Let $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)^T$ be the $n \times p$ data matrix, where each $\mathbf{y}_i$ is a p dimensional vector assumed to be generated from a p-variate normal distribution $N\left(\mathbf{0}, \boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}\right)$. Consider the penalised estimation problem in (2.1) as a departure point for precision matrix estimation. Using an ℓ_2 constraint in place of the ℓ_1 results in a graphical ridge-type problem where the objective is to maximise the log-likelihood

$$\underset{\boldsymbol{\Omega} \in \mathbb{M}^+}{\arg\max} \left\{ \log(\det \boldsymbol{\Omega}) - \operatorname{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right) - \frac{\rho}{2}\|\boldsymbol{\Omega}\|_2^2, \right\}, \tag{3.1}$$

where, $\mathbf{S} = \mathbf{Y}'\mathbf{Y}$ denotes the sample covariance matrix and $\|\boldsymbol{\Omega}\|_2^2 = \sum_{i<j}\sum_{i=1}^{p} \omega_{ij}^2$. The form of (3.1) coincides with the special case of the objective function considered for the type I ridge precision estimator in (Wieringen and Peeters, 2016) and the covariance-regression setting in (Witten and Tibshirani, 2009). The Bayesian connection is given by the maximum MAP estimation of the following model

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$
\begin{aligned}
p\left(\mathbf{y}_i \mid \boldsymbol{\Omega}\right) &= \mathrm{N}_p\left(\mathbf{y}_i \mid \mathbf{0}, \boldsymbol{\Omega}^{-1}\right) \quad (i = 1, \ldots, n) \\
p\left(\boldsymbol{\Omega} \mid \mu = 0, \sigma\right) &= C^{-1} \prod_{i<j} \left\{ \mathrm{N}(\omega_{ij} \mid \sigma) \right\} \prod_{i=1}^{p} \left\{ \mathrm{TN}(\omega_{ii} \mid \sigma) \right\} \mathbb{1}_{\boldsymbol{\Omega} \in \mathbb{M}^+},
\end{aligned}
\tag{3.2}
$$

where, $\mathrm{N}_p(\cdot)$ denotes a p-variate Gaussian probability density function (pdf) with form $p(y) = (\sigma\sqrt{2\pi})^{-1} \exp(-0.5[y/\sigma]^2)$ and $\mathrm{TN}(y \mid \sigma)$ represents a univariate left truncated at zero Gaussian pdf with form $p(y) = \sqrt{2}(\sigma\sqrt{\pi})^{-1} \exp(-0.5[y/\sigma]^2) 1_{y>0}$. Moreover, C is the normalising constant independent of σ given by

$$
C = \int_{\boldsymbol{\Omega} \in \mathbb{M}^+} \prod_{i<j} \left\{ \mathrm{N}(\omega_{ij} \mid \sigma) \right\} \prod_{i=1}^{p} \left\{ \mathrm{TN}(\omega_{ii} \mid \sigma) \right\} d\boldsymbol{\Omega}. \tag{3.3}
$$

Consider the change of variable $\tilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega}/\sigma$, it immediately follows that

$$
\begin{aligned}
C &= \int_{\tilde{\boldsymbol{\Omega}} \in \mathbb{M}^+} \prod_{i<j} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\omega_{ij}^2}{2\sigma^2} \right) \prod_{i=1}^{p} \frac{\sqrt{2}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\omega_{ii}^2}{2\sigma^2} \right) \frac{\partial(\tilde{\omega}_{ij})}{\partial(\omega_{ij})} d\tilde{\boldsymbol{\Omega}} \\
&= \int_{\tilde{\boldsymbol{\Omega}} \in \mathbb{M}^+} \prod_{i<j} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tilde{\omega}_{ij}^2}{2} \right) \prod_{i=1}^{p} \frac{\sqrt{2}}{\sqrt{\pi}} \exp\left(-\frac{\tilde{\omega}_{ii}^2}{2} \right) \frac{\partial(\tilde{\omega}_{ij})}{\partial(\omega_{ij})} d\tilde{\boldsymbol{\Omega}} \\
&= \int_{\tilde{\boldsymbol{\Omega}} \in \mathbb{M}^+} \prod_{i<j} \left\{ \mathrm{N}(\tilde{\omega}_{ij} \mid 1) \right\} \prod_{i=1}^{p} \left\{ \mathrm{TN}(\tilde{\omega}_{ii} \mid 1) \right\} d\tilde{\boldsymbol{\Omega}}.
\end{aligned}
\tag{3.4}
$$

The equality in (3.4) relies on the substitution $\tilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega}/\sigma$, where $g(\tilde{\omega}_{ij}) = \omega_{ij}/\sigma$ for $i \leq j$ and that $\{\tilde{\boldsymbol{\Omega}} : \tilde{\boldsymbol{\Omega}} \in \mathbb{M}^+\} \equiv \{\boldsymbol{\Omega} : \boldsymbol{\Omega} \in \mathbb{M}^+\}$ for $\sigma > 0$. The posterior distribution associated with this prior is given by

$$
\begin{aligned}
p\left(\boldsymbol{\Omega} \mid \mathbf{Y}, \sigma\right) &\propto \prod_{i=1}^{n} \det\boldsymbol{\Omega}^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{y}_i^T \boldsymbol{\Omega} \mathbf{y}_i \right\} \prod_{i<j} \frac{1}{\sqrt{\pi}\sigma} \exp\left(-\frac{\omega_{ij}^2}{2\sigma^2} \right) \prod_{i=1}^{p} \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \exp\left(-\frac{\omega_{ii}^2}{2\sigma^2} \right) \\
&\propto \det\boldsymbol{\Omega}^{\frac{n}{2}} \exp\left\{ -\frac{1}{2} \mathrm{trace}(\mathbf{S}\boldsymbol{\Omega}) - \frac{1}{2\sigma^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \omega_{ij}^2 \right\}.
\end{aligned}
$$

The log of the posterior distribution is

$$
\begin{aligned}
p_\ell\left(\boldsymbol{\Omega} \mid \mathbf{Y}, \sigma\right) &\propto \frac{n}{2} \log(\det\boldsymbol{\Omega}) - \frac{1}{2} \mathrm{trace}\left(\mathbf{S}\boldsymbol{\Omega}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \omega_{ij}^2 \\
&\propto \log(\det\boldsymbol{\Omega}) - \mathrm{trace}\left(\frac{\mathbf{S}}{n} \boldsymbol{\Omega} \right) - \frac{1}{n\sigma^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \omega_{ij}^2.
\end{aligned}
\tag{3.5}
$$

The MAP estimate is the value of $\boldsymbol{\Omega}$ which maximises (3.5)

$$
\begin{aligned}
\hat{\boldsymbol{\Omega}} &= \underset{\boldsymbol{\Omega} \in \mathbb{M}^+}{\arg\max} \left\{ \log(\det\boldsymbol{\Omega}) - \mathrm{trace}\left(\frac{\mathbf{S}}{n} \boldsymbol{\Omega} \right) - \frac{1}{n\sigma^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \omega_{ij}^2 \right\} \\
&= \underset{\boldsymbol{\Omega} \in \mathbb{M}^+}{\arg\max} \left\{ \log(\det\boldsymbol{\Omega}) - \mathrm{trace}\left(\frac{\mathbf{S}}{n} \boldsymbol{\Omega} \right) - \rho_2 \|\boldsymbol{\Omega}\|_2 \right\},
\end{aligned}
$$

where $\rho_2 = 1/n\sigma^2$ and within the Bayesian context σ is treated as the shrinkage parameter.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3.2 Developing a block Gibbs sampler

The block Gibbs sampler described in Section 2.3 of Chapter 2 serves as the impetus for the sampler described here. The target (posterior) distribution associated with the prior in (3.2) is given by:

$$p(\boldsymbol{\Omega} \mid \mathbf{Y}, \sigma) \propto \det\boldsymbol{\Omega}^{\frac{n}{2}} \exp\left\{-\operatorname{trace}\left(\frac{1}{2}\mathbf{S}\boldsymbol{\Omega}\right)\right\} \prod_{i<j}\left\{\exp\left(-\frac{\omega_{ij}^2}{2\sigma^2}\right)\right\} \prod_{i=1}^{p}\left\{\exp\left(-\frac{\omega_{ii}^2}{2\sigma^2}\right)\right\} \mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^+}. \tag{3.6}$$

In contrast to the target distribution in Section 2.3 of Chapter 2 for the Bayesian graphical lasso, the form in (3.6) does not include an unknown latent scale parameter. This light weight formulation may aid in reducing computational speed and complexity, albeit marginally.

Analogous to the block Gibbs sampler described in Section 2.3 of Chapter 2, the sampler here enables posterior computation of $\boldsymbol{\Omega}$ from the target distribution in (3.6). The Gibbs sampler sequentially updates $\boldsymbol{\Omega}$ one column at time. Without loss of generality, consider updating the last column and row of $\boldsymbol{\Omega}$ and let $\boldsymbol{\Upsilon}^R$ be a symmetric $p \times p$ matrix with a zero main diagonal and $\boldsymbol{\tau}^R$ in the upper and lower off-diagonal entries. Here, $\boldsymbol{\tau}^R$ represents a vector of σ^2 values. The exponent 'R' is used purely for styling in order to distinguish from the notation used by the Bayesian graphical lasso Gibbs sampler in Section 2.3 of Chapter 2.

Partition the matrices $\boldsymbol{\Omega}$, $\mathbf{S}$ and $\boldsymbol{\Upsilon}^R$ as follows

$$\underset{p\times p}{\boldsymbol{\Omega}} = \begin{pmatrix} \underset{(p-1)\times(p-1)}{\boldsymbol{\Omega}_{11}} & \underset{(p-1)\times 1}{\boldsymbol{\omega}_{12}} \\ \underset{1\times(p-1)}{\boldsymbol{\omega}_{21}} & \underset{1\times 1}{\omega_{22}} \end{pmatrix}, \quad \underset{p\times p}{\boldsymbol{S}} = \begin{pmatrix} \underset{(p-1)\times(p-1)}{\boldsymbol{S}_{11}} & \underset{(p-1)\times 1}{\boldsymbol{s}_{12}} \\ \underset{1\times(p-1)}{\boldsymbol{s}_{21}} & \underset{1\times 1}{s_{22}} \end{pmatrix}, \quad \underset{p\times p}{\boldsymbol{\Upsilon}^R} = \begin{pmatrix} \underset{(p-1)\times(p-1)}{\boldsymbol{\Upsilon}_{11}^R} & \underset{(p-1)\times 1}{\boldsymbol{\tau}_{12}^R} \\ \underset{1\times(p-1)}{\boldsymbol{\tau}_{21}^R} & \underset{1\times 1}{0} \end{pmatrix}. \tag{3.7}$$

Recall the following

$$\det\boldsymbol{\Omega}^{\frac{n}{2}} = (\omega_{22} - \boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}} \det\boldsymbol{\Omega}_{11}^{\frac{n}{2}} \propto (\omega_{22} - \boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}},$$

since we are only interested in the last column and row. Similarly recall from (2.16),

$$\operatorname{trace}\left(\frac{1}{2}\mathbf{S}\boldsymbol{\Omega}\right) \propto -\frac{1}{2}(2\boldsymbol{s}_{21}\boldsymbol{\omega}_{12} + s_{22}\omega_{22}).$$

The conditional distribution of the last column in (3.6) can be presented as

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$
\begin{aligned}
p(\omega_{12},\omega_{22}\,|\,\boldsymbol{\Omega}_{11},\boldsymbol{\Upsilon}^R,\mathbf{S},\mu) &\propto (\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}}\exp\left(\frac{1}{2}(2\mathbf{s}_{21}\boldsymbol{\omega}_{12}+s_{22}\omega_{22})\right)\\
&\times\exp\left(-\frac{1}{2}\left\{\sum_{i=1}^{p-1}\frac{\omega_{ip}^2}{\sigma^2}+\left(\frac{\omega_{22}}{\sigma}\right)^2\right\}\right)\ (\text{we only focus on }\boldsymbol{\omega_{12}}\text{ and }\omega_{22})\\
&=(\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}}\exp\left(\frac{1}{2}(2\mathbf{s}_{21}\boldsymbol{\omega}_{12}+s_{22}\omega_{22})\right)\\
&\times\exp\left\{-\frac{1}{2}\left(\boldsymbol{\omega}_{12}^{\top}\boldsymbol{D}_{\tau^R}^{-1}\boldsymbol{\omega}_{12}+\left(\frac{\omega_{22}}{\sigma}\right)^2\right)\right\}\\
&=(\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}}\\
&\times\exp\left[-\frac{1}{2}\left\{\boldsymbol{\omega}_{21}\boldsymbol{D}_{\tau^R}^{-1}\boldsymbol{\omega}_{12}+2\boldsymbol{s}_{21}\boldsymbol{\omega}_{12}+\omega_{22}\left(s_{22}+\frac{\omega_{22}}{\sigma^2}\right)\right\}\right]\\
&=\left(\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}\right)^{\frac{n}{2}}\\
&\times\exp\left[-\frac{1}{2}\left\{\boldsymbol{\omega}_{21}\boldsymbol{D}_{\tau^R}^{-1}\boldsymbol{\omega}_{12}+2\boldsymbol{s}_{21}\boldsymbol{\omega}_{12}+\frac{\omega_{22}}{\sigma^2}\left(s_{22}\sigma^2+\omega_{22}\right)\right\}\right]\\
&=\left(\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}\right)^{\frac{n}{2}}\\
&\times\exp\left[-\frac{1}{2}\left\{\boldsymbol{\omega}_{21}\boldsymbol{D}_{\tau^R}^{-1}\boldsymbol{\omega}_{12}+2\boldsymbol{s}_{21}\boldsymbol{\omega}_{12}+\frac{\omega_{22}}{\sigma^2}\left(\sum_{k=0}^{1}\binom{1}{k}(s_{22}\sigma^2)^{1-k}+\omega_{22}^k\right)\right\}\right]\\
&\propto\left(\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}\right)^{\frac{n}{2}}\\
&\times\exp\left[-\frac{1}{2}\left\{\boldsymbol{\omega}_{21}\boldsymbol{D}_{\tau^R}^{-1}\boldsymbol{\omega}_{12}+2\boldsymbol{s}_{21}\boldsymbol{\omega}_{12}+\omega_{22}\left(s_{22}+\frac{1}{\sigma^2}\right)\right\}\right];\ \text{for } k=0,
\end{aligned}
$$

where the penultimate step makes use of the binomial theorem for the expansion of $\left(s_{22}+\frac{\omega_{22}}{\sigma^2}\right)$. Consider the following change of variables

$$
\begin{aligned}
\boldsymbol{\beta} &= \boldsymbol{\omega}_{12}\\
\gamma &= \omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12},
\end{aligned}
$$

with the constant Jacobian not involving $\boldsymbol{\beta}$ nor γ (refer to Section 2.3 in Chapter 2 for more details). This transformation yields the following conditional distribution

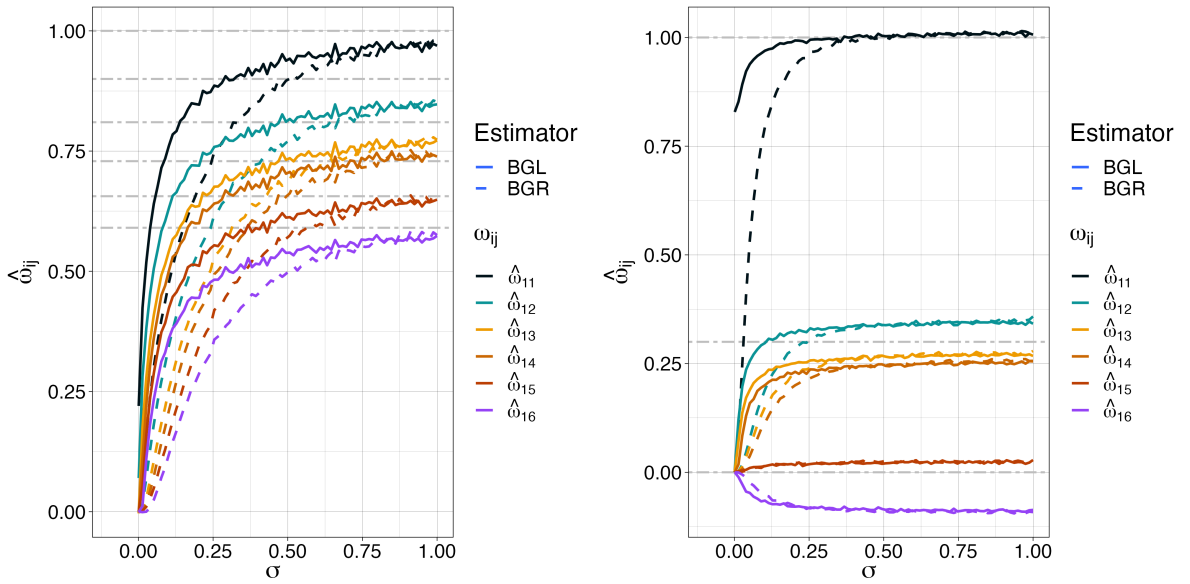
$$
\begin{aligned}
p(\gamma,\boldsymbol{\beta}|\,\boldsymbol{\Omega}_{11},\boldsymbol{\Upsilon}^R,\mathbf{S},\mu) &\propto(\gamma)^{\frac{n}{2}}\exp\left(-\frac{1}{2}\left\{\boldsymbol{\beta}^T\mathbf{D}_{\tau^R}^{-1}\boldsymbol{\beta}+2\mathbf{s}_{21}\boldsymbol{\omega}_{12}+(\gamma+\boldsymbol{\beta}^T\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\beta})\left(s_{22}+\frac{1}{\sigma^2}\right)\right\}\right)\\
&=(\gamma)^{\frac{n}{2}}\exp\left(-\frac{s_{22}+\frac{1}{\sigma^2}}{2}\gamma\right)\exp\left(-\frac{1}{2}\left[\boldsymbol{\beta}^T\left\{\mathbf{D}_{\tau^R}^{-1}+\left(s_{22}+\frac{1}{\sigma^2}\right)\boldsymbol{\Omega}_{11}^{-1}\right\}\boldsymbol{\beta}+2\mathbf{s}_{21}\boldsymbol{\beta}\right]\right).
\end{aligned}
$$

It follows that

$$
(\gamma,\boldsymbol{\beta}|\,\boldsymbol{\Omega}_{11},\boldsymbol{\Upsilon}^R,\mathbf{S},\mu)\sim\text{GA}\left[\frac{n}{2}+1,\left(s_{22}+\frac{1}{\sigma^2}\right)/2\right]\text{N}_{p-1}(-\boldsymbol{C}\boldsymbol{s}_{21},\boldsymbol{C}),\tag{3.8}
$$

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



(a) AR(1) (b) cluster

Fig. 3.2 *Coefficient paths for first row of elements in varying* $\boldsymbol{\Omega}$ *topologies.* Each panel displays the convergence path for the Bayesian graphical lasso (dashed) and the Bayesian graphical ridge (dashed) for $n > p$ and varying σ and $1/\lambda$. The horizontal grey lines (dot-dash) indicate the true values of ω_{ij}.

where $\boldsymbol{C} = \left\{ \left(s_{22} + \frac{1}{\sigma^2} \right) \boldsymbol{\Omega}_{11}^{-1} + \mathbf{D}_{\tau^R}^{-1} \right\}^{-1}$, and $\mathrm{GA}(\cdot)$ refers to a gamma pdf, $\mathrm{N}_{p-1}(\cdot)$ to a $(p-1)$-variate Gaussian pdf and $\mathbf{D}_{\tau^R}^{-1} = \mathrm{diag}(1/\boldsymbol{\tau}_{12}^{\mathrm{R}})$.

Analogous to block Gibbs sampler in Section 2.3 of Chapter 2, the positive definite constraint on $\boldsymbol{\Omega}$ is maintained here due to the archetypal similarities shared by the samplers. The block Gibbs sampler is summarised in Algorithm 3. Conclusively, Figure 3.2 displays the coefficient paths for first row of elements in $\boldsymbol{\Omega}$ for the Bayesian graphical ridge (BGR), as well as the Bayesian graphical lasso (BGL). The 6×6 $\boldsymbol{\Omega}$ follows a topology defined in Section 3.5. The dimensionality is chosen for visual convenience and to avoid redundancy here, the solution paths for the remaining topologies are available in Section B.1 of the appendix.

Algorithm 3 Bayesian graphical ridge: Block Gibbs sampler

Require: Initialise $\boldsymbol{\Omega} \in \mathbb{M}^+$.
for $i = 1, \ldots, p$ **do**
1) Partition $\boldsymbol{\Omega}$, $\mathbf{S}$ and $\boldsymbol{\Upsilon}$ as in (4.7).
2) Sample $\gamma \sim \mathrm{GA}\left(n/2+1, \left[s_{22} + \frac{1}{\sigma^2} \right] /2\right)$ and $\boldsymbol{\beta} \sim \mathrm{N}_{p-1}\left(-\boldsymbol{C}\boldsymbol{s}_{21}, \boldsymbol{C}\right)$.
3) Update $\boldsymbol{\omega}_{12} = \boldsymbol{\beta}$, $\boldsymbol{\omega}_{21} = \boldsymbol{\beta}^\top$ and $\omega_{22} = \gamma + \boldsymbol{\beta}^\top \boldsymbol{\Omega_{11}}^{-1} \boldsymbol{\beta}$.
end for

3.2.1 Computational insights

This section aims to evaluate the performance and scalability of the block Gibbs sampler in Algorithm 3. Here, $\boldsymbol{\Omega}$ is initialised at the identity matrix and, where applicable, synthetic population topologies in Section 3.5 are used. Performance assessments are executed on an Apple silicon (M1 Pro), utilising macOS Ventura (version

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

13.6), with computations being carried out through the R programming language (version 4.2.3). Figure 3.3 displays the amount of time, in minutes, required for 1000 iterations by the Bayesian graphical ridge sampler, where the number of observations n is fixed to 100 and $p \in (25, \cdots, 200)$. Here, a single iteration constitutes the block Gibbs sampler to update each entry of $\mathbf{\Omega}$. For example, the computation time required to generate 1000 iterations when $p = 75$ and $p = 200$ is approximately 37 seconds and 20 minutes respectively. Clearly the block Gibbs samplers have at least $O(p^3)$ complexity, notably the $p^3/3$ floating point operations required by the typical Cholesky decomposition procedure (Golub and Loan, 1996) play a significant role here.

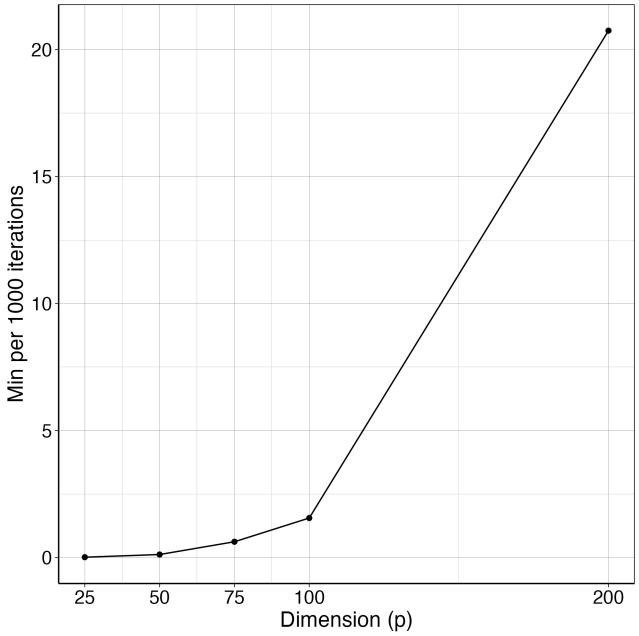


Fig. 3.3 *Bayesian graphical ridge block Gibbs sampler computational insights.* The amount of time, in minutes, required for 1000 iterations from the Bayesian graphical ridge block Gibbs sampler.

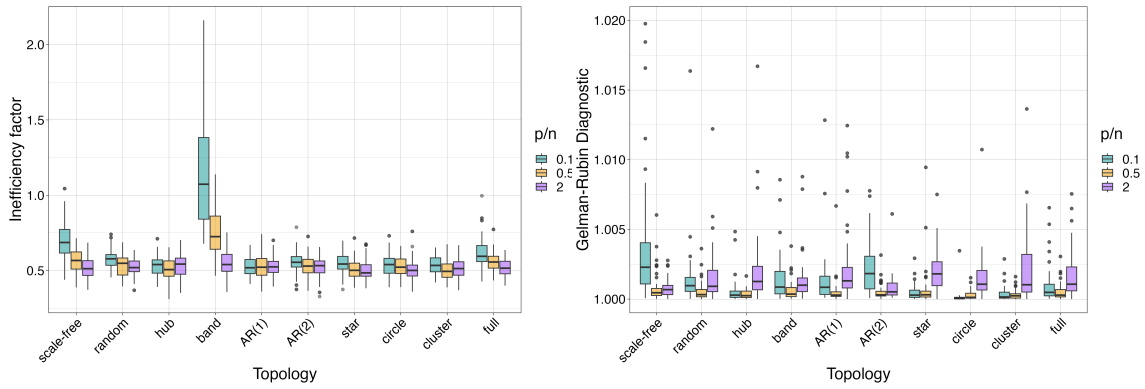
The convergence diagnostics of the Gibbs sampler is assessed using the Gelman-Rubin diagnostic measure of (Gelman and Rubin, 1992) and the inefficiency factor of (Kim et al., 1998) across all elements of the estimated $\mathbf{\Omega}$ and displayed in Figure 3.4. Here the dimension is fixed at $p = 10$ and the sample size varies between $n \in (100, 20, 5)$. The calculation of the inefficiency factor for each network topology (in Section 3.5) involves using 10 000 samples after 5000 burn-in iterations and 1000 lags, of which the median inefficiency factor is calculated. This process is repeated 100 times and the summary results are displayed in Figure 3.4a. Clearly, the inefficiency factor is in the region of 0.5 - 0.7 suggesting minimal information loss due to correlation in the MCMC chains. Finally, Figure 3.4b highlights satisfactory convergence of the Bayesian graphical ridge samplers, with the diagnostic statistic well bellow 1.1, see the detailed discussion in (Vats and Knudson, 2021). Here, the Gelman-Rubin statistic is calculated by using 10 chains for each element of $\mathbf{\Omega}$ and for each network topology.

3.3 Bayesian inference on the precision matrix

This section aims to study the behaviour of relative loss of the Bayesian graphical ridge estimator across varying penalty parameter values, namely σ. In particular, the relative loss of interest here is defined as the ratio of Stein's loss for the Bayesian graphical ridge estimate of $\mathbf{\Omega}$ over Stein's loss of the Bayesian graphical lasso estimate. Recall

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



(a) (b)

Fig. 3.4 *Bayesian graphical ridge block Gibbs sampler convergence diagnostics.* The Bayesian inefficiency factor and the Gelman-Rubin diagnostic across multiple network topologies and p/n ratios are displayed in panel (a) and (b) respectively.

that Bayesian estimators for covariance and precision matrices can be obtained by employing a decision-theoretic approach - see ((Yang and Berger, 1994), (Fouskakis et al., 2009), (Reich, 2012), (Wang, 2012), (Carvalho, 2015), (Kundu et al., 2019)). For example, under Stein's loss function for $\boldsymbol{\Sigma}$

$$L_1\left(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}\right) = \text{trace}\left(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right) - \log\left\{\det\left(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right)\right\} - p,$$

where $\hat{\boldsymbol{\Sigma}}$ is the estimator of $\boldsymbol{\Sigma}$. The Bayes estimator is given by

$$\hat{\boldsymbol{\Sigma}}_{L_1} = E\left(\boldsymbol{\Omega} \,|\, \mathbf{Y}\right)^{-1}. \tag{3.9}$$

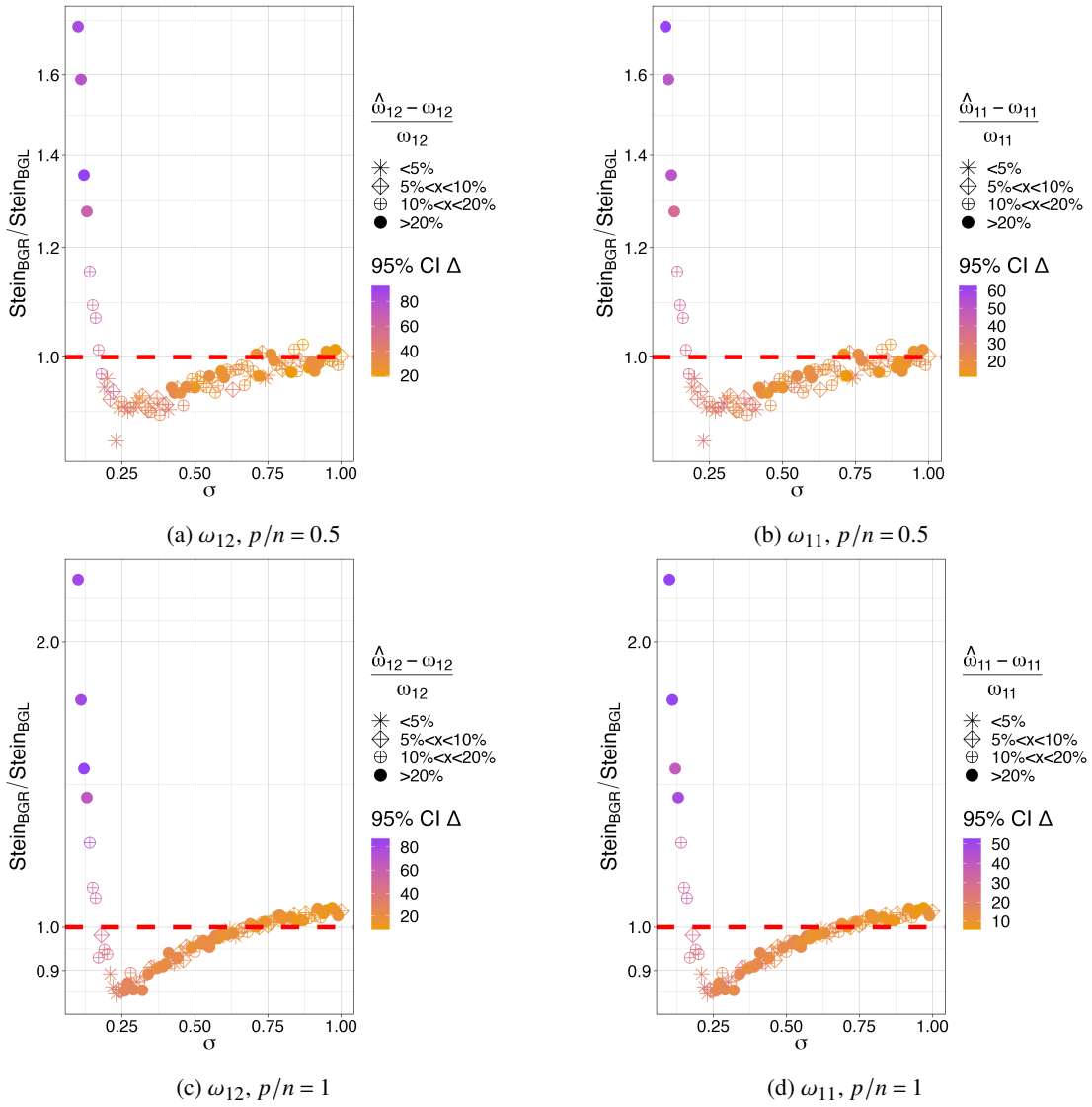
The estimate in (3.9) can be obtained using the the block Gibbs samplers in Section 3.2 and Chapter 2, respectively.

The relative loss of the Bayesian graphical ridge estimator across varying penalty parameter values is also evaluated on data sets drawn from the precision matrices (graphical topologies) defined in Section 3.5. The Bayesian graphical lasso framework was originally designed for moderate dimensions, therefore throughout the simulation, the dimension is fixed at $p = 30$ and the sample size varies between $n \in (60, 30)$. It is worth noting here that Section 3.5 considers scenarios for larger p and $p > n$. In this setting though, for each generated sample, the Bayesian graphical lasso and ridge are fitted across varying σ - the Bayesian graphical lasso uses $\lambda = 1/\sigma$. The Bayesian estimates are based on 10 000 iterations after 5000 burn-in iterations. The ratio of Stein's loss associated with the Bayesian graphical ridge estimator over the Bayesian graphical lasso estimator are approximated using their respective Bayes estimators. Figure 3.5 displays the aforementioned estimates for the full topology. Furthermore, each figure zooms in on the behaviour of a diagonal (ω_{11}) and an off-main diagonal element (ω_{12}). That is, the trend of the plots depict the holistic behaviour of the estimators according to varying penalty parameters and the visual characteristics of each element of the trend highlight the behaviour of individual elements of the estimated $\boldsymbol{\Omega}$. In particular, the colour of each estimate is determined by the width of the standardised credible interval given a particular value of σ. The shape of the estimate reflects the percentage error (since the true value is known). To avoid redundancy here, insights for the remaining topologies, p/n ratios and ω_{11} are provided in Section B.2 of the appendix. A few insightful deductions can be made from this simulation. First, there is a positive association between the degree of uncertainty and the magnitude of the penalty parameter. Smaller values of σ may allow for shrinkage to near zero, however, overshrinkage can negatively impact the accuracy of the precision estimate.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Second, there exists regions of σ wherein the Bayesian graphical ridge outperforms its lasso counterpart, however the size of the region (if present) may vary between topologies. In other words, the selection of the 'correct' σ requires consideration. Moreover, this region is more prominent for larger p/n ratios. Finally, the diagonal and off-diagonal elements associated with the lowest percentage error are typically located in the region of σ where the Stein loss ratio is less than one, i.e. where the Bayesian graphical ridge estimator is the top performer.



(a) ω_{12}, $p/n = 0.5$ (b) ω_{11}, $p/n = 0.5$

(c) ω_{12}, $p/n = 1$ (d) ω_{11}, $p/n = 1$

Fig. 3.5 *Ratio of Stein's loss across a varying penalty parameter for the full network topology*. All panels display, for the full topology, the ratio of Stein's loss across varying σ. The colour gradient highlights the magnitude of the uncertainty associated with ω_{12} for $p/n \in (0.5, 1)$. The shape of the estimate depicts the percentage error associated with the Bayesian graphical ridge estimates.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3.4 Bayesian adaptive graphical ridge

The Bayesian graphical ridge-type prior in (3.2) requires the selection of σ. The structure of the prior and the block Gibbs samplers in Section 3.2 allows for extended hierarchical formulations of the prior to allow for adaptive shrinkage on different off-diagonal elements of $\mathbf{\Omega}$. That is, the Bayesian framework facilitates the usage of a hyperprior on $\kappa = \sigma^2$ and extending the block Gibbs samplers to sample from the posterior distribution thereof. Consider the following Bayesian adaptive graphical ridge-type set of priors

$$\begin{aligned} p\left(\mathbf{y}_i \,|\mathbf{\Omega}\right) &= N_p\left(\mathbf{y}_i \,|\mathbf{0},\, \mathbf{\Omega}^{-1}\right) \quad (i = 1, \cdots, n) \\ p\left(\mathbf{\Omega} \,|\mu = 0\,, \{\kappa_{ij}\}_{i<j}\right) &= C^{-1}_{\{\kappa_{ij}\}_{i<j}} \prod_{i<j} \left\{ \mathrm{N}(\omega_{ij} \,|\, \kappa_{ij}^{\frac{1}{2}}) \right\} \prod_{i=1}^{p} \left\{ \mathrm{TN}(\omega_{ii} \,|\, \kappa_{ii}^{\frac{1}{2}}) \right\} \mathbb{1}_{\mathbf{\Omega} \in \mathbb{M}^+} \\ p\left(\{\kappa_{ij}\}_{i<j} \,|\{\mu_{ii}, \kappa_{ii}\}_{i=1}^{p} \right) &\propto C_{\{\kappa_{ij}\}_{i<j}} \prod_{i<j} \mathrm{IGA}(a, b), \end{aligned} \tag{3.10}$$

where $\mathrm{IGA}(a,b)$ represents an inverse gamma pdf with a shape parameter a and scale parameter b and form $p(y) = b^a/\Gamma(a) y^{-a-1}\exp(-b/y)$. Moreover, $C_{\{\kappa_{ij}\}_{i<j}}$ is the intractable normalisation constant and κ_{ii} are hyperparameters for the main diagonal elements. Notice that the terms $C_{\{\kappa_{ij}\}_{i<j}}$ in (3.10) simplify when sampling from the posterior of κ_{ij}. That being said, model fitting, using the block Gibbs sampler in Algoritm 3 is straight forward by updating κ_{ij} from an inverse gamma distribution - see Algorithm 4.
Conditional on $\mathbf{\Omega}$, the set of posteriors of κ_{ij} is given by

$$\begin{aligned} p\left(\{\kappa_{ij}\}_{i<j} \,|\{\kappa_{ii}\}_{i=1}^{p}, \mathbf{\Omega} \right) &\propto \kappa_{ij}^{-\left(a+\frac{3}{2}\right)} \exp\left(-\frac{\omega_{ij}^2}{2\kappa_{ij}} \right) \exp\left(-\frac{b}{\kappa_{ij}} \right) \\ &= \kappa_{ij}^{-\left(a+\frac{3}{2}\right)} \exp\left(-\frac{1}{\kappa_{ij}} \left\{ b + \frac{\omega_{ij}^2}{2} \right\} \right). \end{aligned}$$

In other words the posterior is again an inverse gamma pdf

$$\kappa_{ij} \,|\, \mathbf{\Omega} \sim \mathrm{IGA}(a + \frac{1}{2}, b + \frac{\omega_{ij}^2}{2}). \tag{3.11}$$

For the selection of a and b in (3.11), a grid search approach is used over all the topologies defined in Section 3.5. For each permutation considered, a and b are determined by the mean values of the lowest 5% of Stein's loss values across all topologies. The estimates used within the calculation of Stein's loss is the the posterior mean estimate of $\mathbf{\Omega}$.

Algorithm 4 Bayesian adaptive graphical ridge: Block Gibbs sampler

Require: Initialise $\mathbf{\Omega} \in \mathbb{M}^+$.
  for $i = 1, \ldots, p$ **do**
    1) Partition $\mathbf{\Omega}$, $\mathbf{S}$ and $\mathbf{\Upsilon}$ as in (4.7).
    2) Sample $\gamma \sim \mathrm{GA}\left(n/2+1, \left[s_{22} + \frac{1}{\kappa_{ii}}\right]/2\right)$ and $\boldsymbol{\beta} \sim \mathrm{N}_{\mathrm{p}-1}\left(-\boldsymbol{C}\boldsymbol{s}_{21}, \boldsymbol{C}\right)$.
    3) Update $\boldsymbol{\omega}_{12} = \boldsymbol{\beta}$, $\boldsymbol{\omega}_{21} = \boldsymbol{\beta}^{\top}$ and $\omega_{22} = \gamma + \boldsymbol{\beta}^{\top} \mathbf{\Omega_{11}}^{-1} \boldsymbol{\beta}$.
  end for
  for $i \neq j$ **do**
    Sample $\kappa_{ij} \sim \mathrm{IGA}(a + \frac{1}{2}, b + \frac{\omega_{ij}^2}{2})$.
  end for

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3.5 Synthetic data analysis

The synthetic study design aims to evaluate the performance of the precision estimates across varying network topologies in terms of both Stein's loss and the Frobenius norm (both are defined below for completeness). Here the dimension $p = 50$ is fixed and the sample size varies $n \in (100, 50, 5)$. The sample sizes are chosen such that the performance of the estimators are evaluated for $n > p$, $n = p$ and $n < p$.

Model configuration. The estimators considered include the Bayesian adaptive graphical lasso in (2.20), the Bayesian adaptive graphical ridge in (3.10), and finally the frequentist alternative type I and II graphical ridge from (Wieringen and Peeters, 2016) as BAGL, BAGR, ALTI and ALTII, respectively. The parameters of the gamma priors in the BAGL estimator are set to $r = 10^{-2}$ and $s = 10^{-6}$, as proposed by the founding author. Additionally, the parameters of the inverse gamma priors for the BAGR estimator are set to $a = 1$ and $b = 10^{-2}$. The penalty parameter of the ALTI and ALTII are selected using the founding author's automatic optimal penalty search function from their R package 'rags2ridges'. Moreover, the target matrices for the frequentist estimators are given by a null and diag($\mathbf{S}$). The choice of the target matrices is based on the simulations considered by the original authors. The Bayesian samplers are implemented using the 'baygel' R package and are configured to use 10 000 iterations after 5000 burn-in runs. The basic assumptions for all simulations is that $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ are generated from a $N_p(0, \boldsymbol{\Sigma})$ where the true precision is given by $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. To assist the reading experience, Table 3.1 provides a summary of the algorithms considered in this study.

Table 3.1 *Summary of the precision matrix algorithms considered in the synthetic data analysis study.*

Abbreviation	Full name	Author
BAGL	Bayesian adaptive graphical lasso	(Wang, 2012)
BAGR	Bayesian adaptive graphical ridge	(Smith et al., 2022a)
ALTI	type I graphical ridge	(Wieringen and Peeters, 2016)
ALTII	type II graphical ridge	(Wieringen and Peeters, 2016)

Simulation configuration. The population precision matrices considered are intended to explore the performance of the estimators across varying degrees of sparsity and are given by:

1. $\boldsymbol{\Omega}^{\text{AR}(1)}$ with elements given by $\omega_{ij} = 0.7^{|i-j|}$. Here, AR(1) refers to a first-order autoregressive process.
2. $\boldsymbol{\Omega}^{\text{cluster}(S_1:S_2)}$ with elements given by $\omega_{ii} = 1, \omega_{ij} = 0.3$ for $1 \leq i \neq j \leq S_1 < p$, $\omega_{ij} = 0.3$ for $S_1 + 1 \leq i \neq j \leq p$ and $\omega_{ij} = 0$ otherwise. Here, the precision matrix is composed of two hubs, the first hub occupies rows and columns $1 : S_1$ and is of size $S_1 \times S_1$. The second hub occupies rows and columns $(S_1 + 1) : p$ and is of size $S_2 \times S_2$ where $S_2 = (p - S_1)$. The elements of $\boldsymbol{\Omega}$ not occupied by the hubs are set to zero.
3. $\boldsymbol{\Omega}^{\text{star}}$ with elements given by $\omega_{ii} = 1, \omega_{1,j+1} = 0.1 = \omega_{j+1,1}$ for $j = 1, \cdots, p - 1$ and zero otherwise.
4. $\boldsymbol{\Omega}^{\text{full}}$ with elements given by $\omega_{ii} = 1$ and $\omega_{ij} = 0.5$.
5. $\boldsymbol{\Omega}^{\text{scale-free}}$ with elements generated using the B-A algorithm via the 'huge' R package. In this case, the graph attribute is set to 'scale-free'. All remaining default configuration is used besides n and p.
6. $\boldsymbol{\Omega}^{\text{random}}$ with elements generated using the 'huge' R package. In this case, the graph attribute is set to 'random'. All remaining default configuration is used besides n and p.
7. $\boldsymbol{\Omega}^{\text{hub}}$ with elements generated using the 'huge' R package. In this case, the graph attribute is set to 'hub'. All remaining default configuration is used besides n and p.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

8. $\boldsymbol{\Omega}^{\text{band}}$ with elements generated using the 'band' R package. In this case, the graph attribute is set to 'random'. All remaining default configuration is used besides n and p.

9. $\boldsymbol{\Omega}^{\text{AR}(2)}$ with elements given by $\omega_{ii} = 1, \omega_{i,i-1} = \omega_{i-1,i} = 0.5$ and $\omega_{i,i-2} = \omega_{i-2,i} = 0.25$ and 0 otherwise. Here, AR(2) refers to a second-order autoregressive process.

10. $\boldsymbol{\Omega}^{\text{circle}}$ with elements given by $\omega_{ii} = 1, \omega_{i,i-1} = \omega_{i-1,i} = 0.5$ and $\omega_{1,p} = \omega_{p,1} = 0.1$.

Loss functions. The loss functions used within this simulation are given by

- Stein's loss: $\text{trace}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log(|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}|) - p$.
- Frobenius norm: $\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\|_F$, where $\|\boldsymbol{A}\|_F = \sqrt{\text{trace}(\boldsymbol{A}\boldsymbol{A}^\top)}$ and referred to as F norm within the study.

Simulation results. The results presented in Tables 3.2, 3.3, and 3.4 offer valuable insights into the behaviour of the estimation techniques considered. The loss values for the Bayesian estimators are calculated using the posterior mean estimates. The values presented in the aforementioned tables are calculated as the median of the loss values by repeating the estimation procedure 30 times. The standard errors of the estimates are also provided.

When the ratio of variables to samples (p/n) is 0.5, the BAGL estimator stands out as the top performer for sparser network topologies. The BAGR estimator showcases strong performance under dense topolgies such as the AR(1), cluster and the full types. Under both sparsity configurations, the BAGR estimator generally being the runner up. For $p/n = 1$, the BAGR and BAGL estimator enjoy near equal success across the various sparse topologies, whereas the the former takes podium position for dense topplogies.

In the high $p/n = 10$ scenario, BAGR is the clear leader across all topologies under Stein's loss and the F norm. It is also worth mentioning that, in this setting, the Bayesian ridge estimator generally enjoys the lower standard errors.

Overall, these findings provide valuable guidance on the performance of the estimation techniques across various network topologies and p/n ratios.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3.2 *Summary of the F norm and Stein's loss across multiple network topologies and estimators for* $p/n = 0.5$. The median loss values and (standard errors) reported here are based on 30 replications for the BAGL, BAGR, ALTI and ALTII estimators. The Bayesian estimates are obtained as the mean of the posterior distribution. The cell colour of the best performing values are set to grey.

$\frac{p}{n} = 0.5$	**AR(1)**				**AR(2)**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	8.75	11.1	23.0	22.1	3.93	6.92	11.9	12.1
	(0.64)	(0.31)	(1.13)	(0.87)	(0.40)	(0.29)	(0.50)	(0.53)
F norm	7.09	6.63	9.04	8.88	2.67	3.54	4.80	4.56
	(0.20)	(0.07)	(0.10)	(0.08)	(0.22)	(0.11)	(0.09)	(0.07)
$\frac{p}{n} = 0.5$	**star**				**circle**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	2.29	4.27	1.28	7.07	2.24	6.60	10.9	10.9
	(0.19)	(0.14)	(0.13)	(0.39)	(0.25)	(0.35)	(0.52)	(0.53)
F norm	2.56	4.00	1.44	2.48	1.53	3.37	3.52	3.52
	(0.16)	(0.09)	(0.07)	(0.03)	(0.17)	(0.14)	(0.10)	(0.10)
$\frac{p}{n} = 0.5$	**scale-free**				**random**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	2.88	4.32	4.45	8.53	3.65	4.60	6.21	8.89
	(0.28)	(0.20)	(0.36)	(0.50)	(0.36)	(0.17)	(0.51)	(0.46)
F norm	3.24	4.20	4.05	4.25	3.74	4.10	4.41	4.41
	(0.21)	(0.18)	(0.18)	(0.16)	(0.22)	(0.10)	(0.23)	(0.17)
$\frac{p}{n} = 0.5$	**cluster**				**hub**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	13.6	11.4	17.3	17.4	2.88	4.35	6.39	9.49
	(0.43)	(0.18)	(0.66)	(0.64)	(0.37)	(0.24)	(0.35)	(0.45)
F norm	14.2	13.9	14.4	14.2	3.54	4.87	5.89	5.83
	(0.03)	(0.02)	(0.02)	(0.02)	(0.31)	(0.19)	(0.07)	(0.05)
$\frac{p}{n} = 0.5$	**band**				**full**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	2.48	4.20	7.86	9.53	33.2	20.0	44.9	35.8
	(0.27)	(0.19)	(0.42)	(0.55)	(1.06)	(0.26)	(1.12)	(1.13)
F norm	3.14	3.78	4.93	4.91	24.8	24.1	25.0	24.8
	(0.27)	(0.10)	(0.15)	(0.13)	(0.02)	(0.01)	(0.01)	(0.02)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3.3 *Summary of the F norm and Stein's loss across multiple network topologies and estimators for* $p/n = 1$. The median loss values and (standard errors) reported here are based on 30 replications for the BAGL, BAGR, ALTI and ALTII estimators. The Bayesian estimates are obtained as the mean of the posterior distribution. The cell colour of the best performing values are set to grey.

$\frac{p}{n} = 1$	**AR(1)**				**AR(2)**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	15.1	17.6	52.5	42.4	8.37	10.7	20.5	21.3
	(1.63)	(0.54)	(5.34)	(2.84)	(0.58)	(0.33)	(1.31)	(1.22)
F norm	7.99	7.80	10.3	9.89	4.20	4.65	5.88	5.48
	(16.9)	(0.08)	(0.13)	(0.10)	(6.43)	(0.11)	(0.14)	(0.10)
$\frac{p}{n} = 1$	**star**				**circle**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	4.32	5.23	1.95	11.1	19.2	10.1	20.5	20.6
	(0.49)	(0.24)	(0.24)	(0.75)	(13.7)	(0.55)	(1.32)	(1.34)
F norm	4.41	4.65	1.83	2.71	45.1	3.87	4.53	4.53
	(0.44)	(0.17)	(0.17)	(0.03)	(15)	(0.16)	(0.15)	(0.15)
$\frac{p}{n} = 1$	**scale-free**				**random**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	5.30	5.51	5.44	13.2	6.19	5.85	7.78	14.0
	(0.41)	(0.24)	(0.63)	(0.89)	(0.43)	(0.27)	(1.04)	(1.02)
F norm	5.05	4.90	4.43	4.72	5.48	4.68	4.85	5.05
	(0.39)	(0.29)	(0.22)	(0.18)	(0.39)	(0.15)	(0.33)	(0.25)
$\frac{p}{n} = 1$	**cluster**				**hub**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	14.6	14.0	18.2	24.0	5.42	5.62	7.95	14.6
	(0.58)	(0.25)	(0.74)	(1.31)	(0.51)	(0.23)	(0.84)	(0.90)
F norm	14.3	14.3	14.3	14.3	5.40	5.67	6.15	6.23
	(0.06)	(0.04)	(0.02)	(0.03)	(0.45)	(0.12)	(0.14)	(0.07)
$\frac{p}{n} = 1$	**band**				**full**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	5.76	6.21	11.2	15.5	31.7	24.3	45.6	45.7
	(0.42)	(0.21)	(0.49)	(0.79)	(1.25)	(0.28)	(1.45)	(2.27)
F norm	5.57	4.57	5.88	5.76	24.7	24.4	25.0	24.9
	(0.35)	(0.10)	(0.10)	(0.12)	(0.03)	(0.02)	(0.01)	(0.02)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3.4 *Summary of the F norm and Stein's loss across multiple network topologies and estimators for* $p/n = 10$. The median loss values and (standard errors) reported here are based on 30 replications for the BAGL, BAGR, ALTI and ALTII estimators. The Bayesian estimates are obtained as the mean of the posterior distribution. The cell colour of the best performing values are set to grey.

$\frac{p}{n} = 10$	**AR(1)**				**AR(2)**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	59.6	33.3	85.2	103	45.3	17.9	38.4	51.5
	(1.33)	(0.26)	(9.20)	(11.1)	(1.23)	(0.30)	(5.82)	(6.87)
F norm	29.3	9.92	11.0	10.7	28.0	6.27	10.1	6.35
	(1.24)	(0.03)	(0.96)	(0.10)	(1.17)	(0.07)	(9.87)	(0.22)
$\frac{p}{n} = 10$	**star**				**circle**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	33.5	4.23	14.2	27.9	50.2	26.2	29.2	27.6
	(1.18)	(0.10)	(2.85)	(4.04)	(2.52)	(1.01)	(53.0)	(62.1)
F norm	28.0	3.68	13.9	2.06	29.5	5.25	7.68	7.73
	(1.08)	(0.05)	(9.60)	(0.17)	(2.51)	(0.12)	(0.44)	(0.19)
$\frac{p}{n} = 10$	**scale-free**				**random**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	30.6	5.68	18.0	33.8	29.1	7.19	21.9	36.7
	(1.23)	(0.28)	(4.33)	(4.18)	(1.45)	(0.46)	(3.95)	(3.65)
F norm	27.1	4.64	15.1	4.64	26.0	5.00	16.5	5.70
	(1.29)	(0.12)	(21.1)	(0.22)	(1.42)	(0.22)	(9.85)	(0.31)
$\frac{p}{n} = 10$	**cluster**				**hub**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	44.1	16.9	29.1	44.3	30.2	6.96	20.3	35.2
	(1.56)	(0.18)	(2.72)	(4.19)	(1.32)	(0.25)	(3.46)	(3.94)
F norm	30.5	14.6	16.9	14.4	26.8	6.11	15.2	6.44
	(1.36)	(0.02)	(4.38)	(0.05)	(1.25)	(0.08)	(12.1)	(0.18)
$\frac{p}{n} = 10$	**band**				**full**			
	BAGL	BAGR	ALTI	ALTII	BAGL	BAGR	ALTI	ALTII
Stein	28.6	9.42	25.7	39.0	59.1	32.3	56.7	72.5
	(1.14)	(0.20)	(4.24)	(4.85)	(1.39)	(0.21)	(4.58)	(6.99)
F norm	25.4	5.69	16	6.58	36.7	24.9	26.0	25.0
	(1.26)	(0.04)	(17.8)	(0.33)	(1.11)	(0.01)	(10)	(0.04)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3.6 Bayesian structure learning

This section briefly addresses structure learning options for the Bayesian estimators evaluated in Section 3.5, followed by an application on the flow cytometry dataset from (Sachs et al., 2005). The continuous priors of the Bayesian estimators considered require a heuristic treatment for edge selection. Three treatments are investigated here: the decision-theoretic thresholding approach as defined in (2.11) of Chapter 2, denoted as TAI; next the absolute value of the ratio in TAI, denoted as TAII is also considered. The motivation for the use of TAII is to explore the performance of the decision-theoretic approach of TAI but where it is independent of whether the association is positive or negative; it is primarily influenced by the magnitude of the estimate. Additionally, similar to the Bayesian graphical horseshoe estimator in (Li et al., 2019), the symmetric central 50%, 90% and 99% posterior credible intervals are utilised as an analytical approach. These methods are named 50CI, 90CI and 99CI respectively. Here an edge is selected if 0 is not in the credible interval. To assist the reading experience, Table 3.5 provides a summary of the heuristic structure learning techniques considered in this study.

Table 3.5 *Summary of the heuristic structure learning techniques considered in the synthetic data analysis study.*

Abbreviation	Description	Edge selection criterion
TAI	$\tilde{\rho}_{ij}/E_g(\rho_{ij}\,\vert\,\mathbf{Y})$	>0.5
TAII	$\lvert\tilde{\rho}_{ij}/E_g(\rho_{ij}\,\vert\,\mathbf{Y})\rvert$	>0.5
50CI	50% symmetric credible interval for each $\hat{\omega}_{ij}\in\hat{\mathbf{\Omega}}$	0 not in credible interval
90CI	90% symmetric credible interval for each $\hat{\omega}_{ij}\in\hat{\mathbf{\Omega}}$	0 not in credible interval
95CI	95% symmetric credible interval for each $\hat{\omega}_{ij}\in\hat{\mathbf{\Omega}}$	0 not in credible interval

Performance functions. Only the BAGL and BAGR estimators are considered in the evaluation, with assessment based on

- True positive (TP): an edge is correctly predicted.
- False positive (FP): an edge is incorrectly predicted.
- True negative (TN): the absence of an edge is correctly predicted.
- False negative (FN): the absence of an edge is incorrectly predicted.
- Sensitivity: $\frac{\text{TP}}{\text{TP+FN}}$. The sensitivity, also referred to as the true positive rate, measures the ability of the structure learning technique to correctly classify an edge.
- Specificity: $\frac{\text{TN}}{\text{TN+FP}}$. The specificity, also referred to as the true negative rate, measures the ability of the structure learning technique to correctly classify the absence of an edge.
- F1 score: $\frac{2\text{TP}}{2\text{TP+FN+FP}}$. The F1 score measures the overall accuracy of the structure learning technique and is the harmonic mean of precision (true edges divided by the number of all predicted edges) and sensitivity.
- Matthews correlation coefficient (MCC): $\frac{\text{TP}\times\text{TN}-\text{FP}\times\text{FN}}{\sqrt{(\text{TP+FP})(\text{TP+FN})(\text{TN+FP})(\text{TNFN})}}$. The MCC statistical rate is more reliable and yields a high score only when the prediction performs well in all four categories of the confusion matrix (Chicco and Jurman, 2020).

Higher values indicate superior performance for all metrics. The simulation configuration as in Section 3.5 is also adopted here. Tables 3.7 and 3.6 summarise the sensitivity, specificity, F1 and MCC scores across multiple network

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

topologies, Bayesian estimators and edge selection techniques TAI, TAII, 50CI, 90CI and 99CI respectively for $p/n = 0.5$. To avoid redundancy here, the results for $p/n = 1$ and $p/n = 10$ are available in Subsection B.3 in the appendix.

Simulation results. Perusing the results it is clear that the BAGL estimator enjoys better performance across sparser topologies whilst BAGR the denser for $n < p$. In this setting the credible interval edge selection techniques are clearly superior. The 90CI and 99CI edge selection technique are generally the top performers across all metrics for both BAGR and BAGL. Exceptions to this pattern are present, however, for instance under the setting where $n = p$ the BAGR and BAGL display similar performance, where the former slightly outperforms the latter, and the best performing edge selection technique varies between the confidence threshold candidates. These results are similar when $n < p$, however, under this setting the Wishart thresholding techniques prove useful under the denser and some scarce topologies. Reviewing the results holistically, the specificity of the dense toplogies are unsurprisingly low due the absence of of true negatives. The cluster topology proves challenging for both Bayesian estimators under all edge selection techniques in terms of adequately minimising false positives. It is clear that for all simulations considered here, the analytical credible interval edge selection techniques generally outperform the Wishart thresholding approaches, however, the performance of all techniques considered here rely on the accuracy of the underlying precision matrix estimator.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3.6 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 0.5$. The median loss values reported here are based on 30 replications for the BAGL estimator. The cell colour of the best performer is set to grey.

$\frac{p}{n} = 0.5$	**AR(1)**					**AR(2)**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.98	1	1	0.96	0.95	0.7	0.92	0.96	0.64	0.53
Specificity	0.11	0.1	0.1	0.1	0.1	0.99	0.98	0.97	0.98	0.98
F1	0.23	0.17	0.15	0.19	0.23	0.78	0.85	0.83	0.72	0.64
MCC	0.09	0.1	0.09	0.07	0.06	0.76	0.84	0.83	0.69	0.62
$\frac{p}{n} = 0.5$	**star**					**circle**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.41	0.79	0.9	0.31	0.25	0.77	0.97	0.99	0.78	0.66
Specificity	0.97	0.96	0.96	0.96	0.96	1	1	1	1	1
F1	0.44	0.5	0.51	0.36	0.33	0.86	0.98	0.99	0.87	0.79
MCC	0.4	0.52	0.55	0.32	0.29	0.86	0.98	0.99	0.87	0.79
$\frac{p}{n} = 0.5$	**scale-free**					**random**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.52	0.85	0.93	0.43	0.34	0.63	0.9	0.96	0.53	0.41
Specificity	0.98	0.98	0.97	0.98	0.98	0.98	0.97	0.96	0.98	0.98
F1	0.61	0.7	0.68	0.52	0.45	0.7	0.74	0.69	0.6	0.53
MCC	0.59	0.7	0.69	0.5	0.43	0.67	0.73	0.71	0.56	0.5
$\frac{p}{n} = 0.5$	**cluster**					**hub**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.94	0.97	1	0.85	0.86	0.56	0.85	0.92	0.44	0.36
Specificity	0.08	0.08	0.08	0.07	0.07	0.99	0.98	0.98	0.99	0.99
F1	0.12	0.05	0.05	0.06	0.1	0.67	0.8	0.79	0.57	0.5
MCC	0.02	0.03	0.04	0	0	0.67	0.79	0.79	0.57	0.51
$\frac{p}{n} = 0.5$	**band**					**full**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.65	0.92	0.96	0.57	0.45	1	1	1	1	1
Specificity	1	1	1	1	1	0	0	0	0	0
F1	0.78	0.94	0.95	0.72	0.62	0.12	0.05	0.04	0.04	0.05
MCC	0.79	0.93	0.95	0.72	0.64	0.01	0	0	0	0

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3.7 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 0.5$. The median loss values reported here are based on 30 replications for the BAGR estimator. The cell colour of the best performer is set to grey.

$\frac{p}{n} = 0.5$	**AR(1)**					**AR(2)**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.91	0.97	1	0.92	0.9	0.23	0.7	0.94	0.27	0.2
Specificity	0.11	0.11	0.1	0.1	0.1	0.99	0.98	0.97	0.96	0.96
F1	0.61	0.24	0.16	0.39	0.5	0.37	0.75	0.79	0.39	0.32
MCC	0.03	0.08	0.09	0.02	0	0.37	0.72	0.78	0.35	0.26
$\frac{p}{n} = 0.5$	**star**					**circle**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.13	0.51	0.9	0.15	0.13	0.16	0.64	0.91	0.26	0.19
Specificity	0.98	0.97	0.96	0.97	0.97	1	1	1	1	1
F1	0.22	0.47	0.51	0.23	0.21	0.27	0.78	0.95	0.41	0.32
MCC	0.21	0.44	0.55	0.2	0.19	0.32	0.78	0.95	0.46	0.37
$\frac{p}{n} = 0.5$	**scale-free**					**random**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.16	0.66	0.94	0.18	0.14	0.22	0.77	0.96	0.24	0.18
Specificity	1	0.98	0.97	0.98	0.98	1	0.98	0.96	0.97	0.97
F1	0.28	0.68	0.61	0.29	0.23	0.35	0.75	0.64	0.35	0.29
MCC	0.32	0.66	0.64	0.27	0.22	0.38	0.72	0.66	0.32	0.25
$\frac{p}{n} = 0.5$	**cluster**					**hub**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.93	0.96	1	0.69	0.79	0.16	0.66	0.9	0.19	0.14
Specificity	0.08	0.08	0.08	0.08	0.08	1	0.98	0.97	0.98	0.98
F1	0.58	0.14	0.06	0.08	0.2	0.28	0.69	0.63	0.29	0.24
MCC	0.03	0.04	0.05	0	0	0.33	0.67	0.65	0.29	0.24
$\frac{p}{n} = 0.5$	**band**					**full**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.19	0.84	0.99	0.27	0.2	1	1	1	1	1
Specificity	1	1	0.99	0.99	0.99	0	0	0	0	0
F1	0.32	0.91	0.93	0.41	0.33	0.65	0.2	0.08	0.04	0.08
MCC	0.38	0.91	0.93	0.45	0.37	0.02	0.01	0	0	0

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3.6.1 Cell signalling analysis

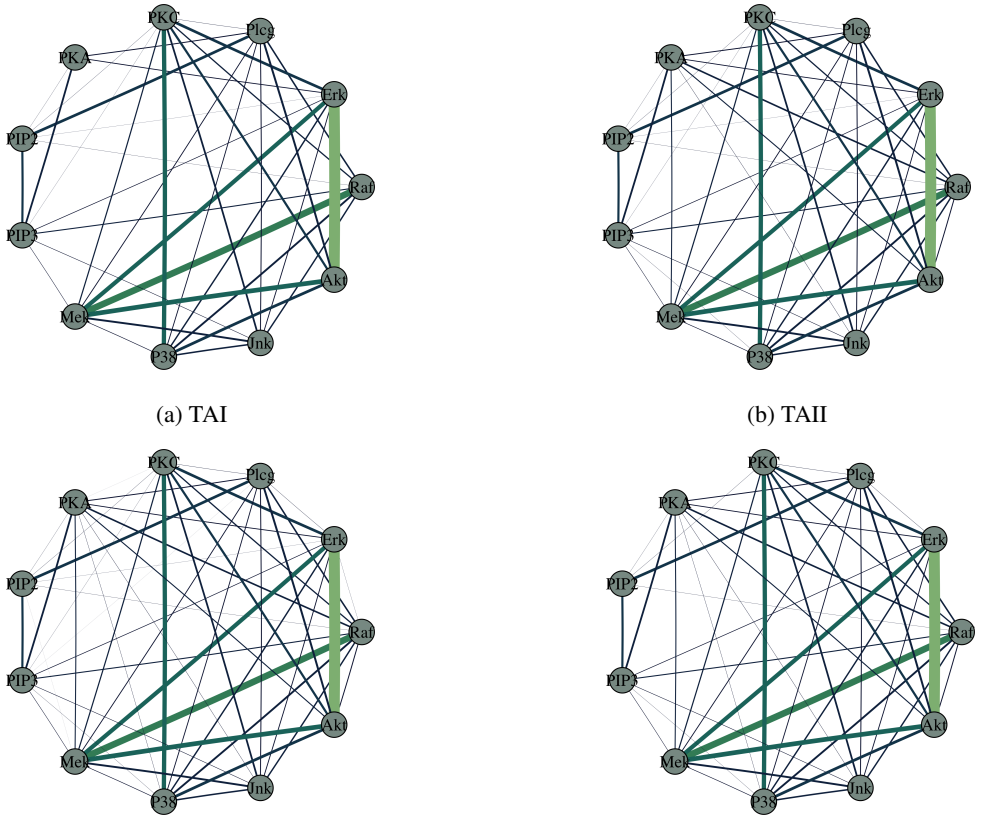
The goal of the analysis is to provide inference on the signalling network that connects key phosphorylated proteins in human T cell signalling. The data consist of $p = 11$ proteins and $n = 7466$ cells. Flow cytometry data sets typically present thousands of individual cells worth of independent observations making it an ideal application for ridge based modelling. The BAGR estimates are based on the MCMC outputs consisting of 10 000 iterations after 5000 burn-ins. Figure 3.6 displays the inferred networks by the BAGR estimator. The width of the edges represent the strength of the association derived from the estimated precision matrix and the presence of an edge is determined by the TAI, TAII, 50CI and 99CI edge selection technique. To evaluate the validity of these estimates, Table 3.8 highlights which of the edge selection techniques are cable of detecting edges well-established in literature - refer to the supplementary material of (Sachs et al., 2005) for the list of citations. The thresholding approaches produce the sparsest networks, however, at the cost of well established edges. In contrast, the credible interval approaches produce denser networks, and unsurprisingly, capture the edges of interest. The structure learning results from the synthetic studies performed in Section 3.6 motivate the use of credible intervals for edge selection. Furthermore, experiments in (Friedman et al., 2008) suggest using the unregularised model based on the lowest cross validation error obtained in search of the optimal shrinkage parameter. These results support the networks produced by the BAGR estimator, however, there may exist edges, not listed in Table 3.8 but present in Figure 3.6, that require further investigation by biologists to distinguish information from noise. Moreover, the concentration between edges or signalling intensities, may provide valuable insights into the extent to which one protein's activity influences another, see for example (Shin et al., 2011).

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3.8 *Evaluation of established molecule edges by BAGR and BAGL. Refer to Table 3 in (Sachs et al., 2005).*

Connection	Edge selection technique			
	TAI	TAII	50CI	99CI
PKC - Raf	✓	✓	✓	✓
PKC - Mek	✓	✓	✓	✓
PKC - Jnk	✓	✓	✓	✓
PKC - p38	✓	✓	✓	✓
PKC - PKA	×	×	✓	×
PKA - Raf	×	✓	✓	✓
PKA - Mek	×	✓	✓	✓
PKA - Erk	✓	✓	✓	✓
PKA - Jnk	×	✓	✓	✓
PKA - p38	×	×	✓	✓
Raf - Mek	✓	✓	✓	✓
PKA - Akt	×	✓	✓	✓
Mek - Erk	✓	✓	✓	✓
Plcg - PIP_2	✓	✓	✓	✓
Plcg - PIP_3	×	×	✓	×
PIP_3 - PIP_2	✓	✓	✓	✓
Erk - Akt	✓	✓	✓	✓



(a) TAI

(b) TAII

(c) 50CI

(d) 99CI

Fig. 3.6 *Undirected cell-signalling graphs estimated using the BAGR estimator and various edge selection techniques.*

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3.7 Conclusion

The Bayesian graphical ridge-type prior and its adaptive variant is introduced, for the low to moderate dimension setting, along with its block Gibbs sampler. This Bayesian estimator rooted in absolute continuous priors is shown to be attractive for estimating precision matrices under the assumption of Gaussianity. The new estimator displays commendable performance when compared to the frequentist alternative type I and II estimators and out performs the Bayesian graphical lasso for denser topologies when $n \leq p$, and when $n > p$ across all topologies considered. The computational cost of the Bayesian graphical ridge-type estimators is reasonable in the low to moderate dimension setting but may struggle in the high-dimensional scenario. Juxtaposed to the common comparison to frequentist methods, this study instead demonstrates the plausible inferential capabilities due to the richness of the information contained in the posterior distribution. For example, in structure learning or measuring uncertainty related to the magnitude of the shrinkage quantity. A heuristic Bayesian structure learning comparison is briefly evaluated, namely a partial correlation thresholding and symmetric credible interval approach. The credible interval approach generally yields the best performing metrics. To this end, the Bayesian adaptive graphical ridge-type estimator demonstrates the ability to successfully infer well established associations between key phosphorylated proteins in human T cell signalling that may provide valuable clinical inference. The Bayesian graphical ridge estimator employed in this chapter is implemented in the 'baygel' R-package which is freely available from the Comprehensive R Archive Network.

Finally, whilst the Bayesian graphical ridge method performs well on denser network topologies for $n \leq p$ and in general for $n > p$, practical scenarios may lack prior knowledge of true sparsity levels and the volume of available data. Researchers could benefit from an adaptive technique that combines the strengths of both lasso and ridge approaches, ensuring effectiveness across varying degrees of sparsity and p/n ratios in different topologies. This is exactly, the focus of the next chapter.

© University of Pretoria

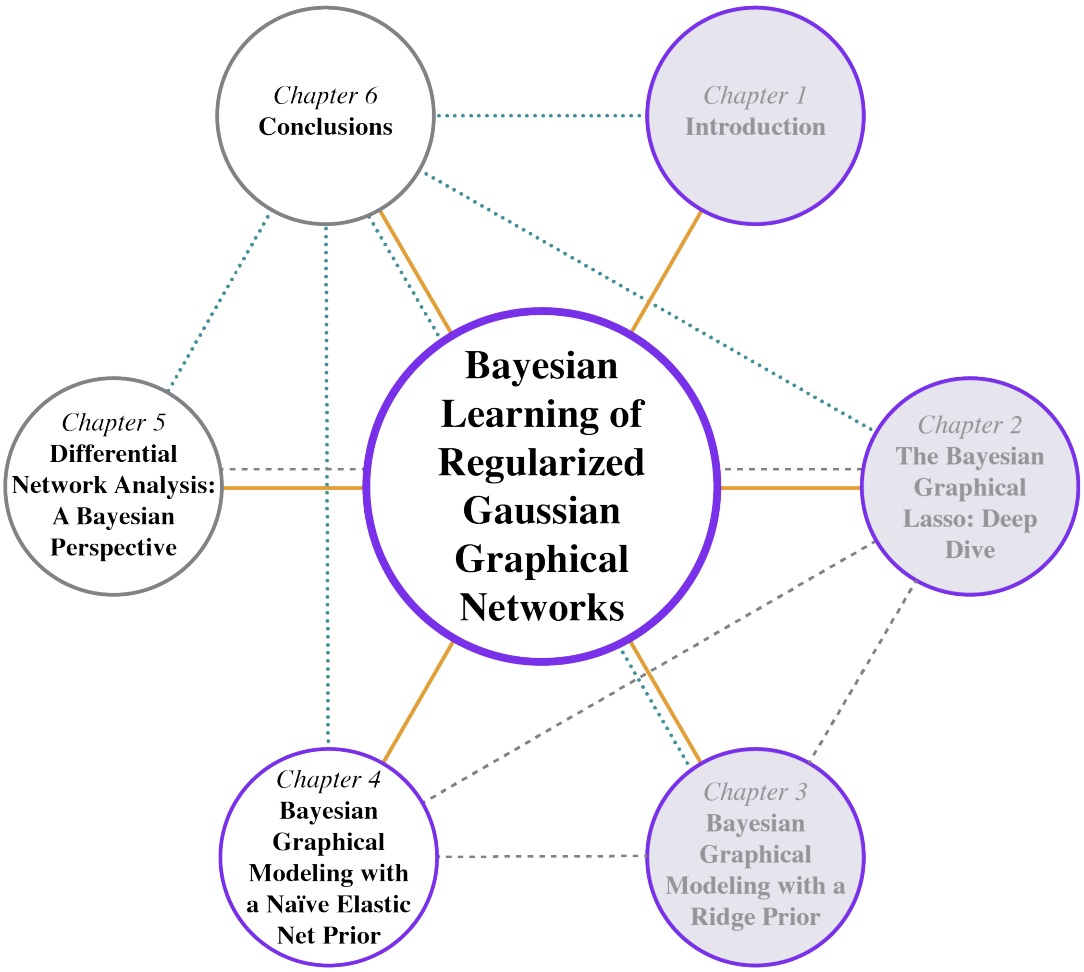
© University of Pretoria

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chapter 4

Bayesian Graphical Modeling with a Naïve Elastic Net Prior



© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

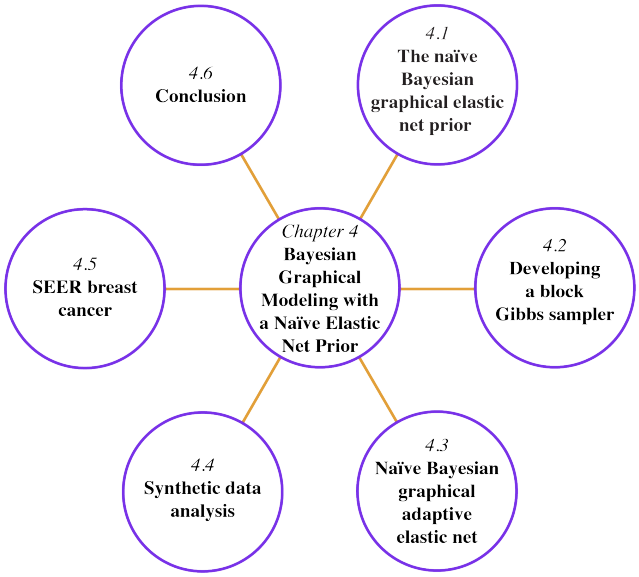


Fig. 4.1 *Chapter 4 visual outline.*

In scenarios with sparse precision matrices and ample data, the Bayesian graphical lasso excels, while the Bayesian graphical ridge prioritises accuracy over sparsity, especially when variables outnumber observations. However, challenges arise when estimating without prior knowledge of sparsity. This chapter introduces a novel Bayesian estimator for the inverse covariance matrix under the assumption of Gaussianity. The new estimator merges the strengths of previous methods, aiming for versatility in addressing various network structures and data-scarce situations. Figure 4.1 provides a visual outline of this chapter.

4.1 The naïve Bayesian graphical elastic net prior

Considering the extensive array of possible network topologies, each with its unique degree of sparsity and precision magnitude, selecting the most suitable estimation technique can be a complex task. The heterogeneity inherent in these topologies necessitates an adaptable estimation approach, capable of assimilating strengths from various techniques, see a frequentist approach in (Shutta et al., 2023). Considering the unique advantages offered by the Bayesian graphical lasso in Chapter 2 and ridge in Chapter 3, the logical progression is to merge these shrinkage techniques within the framework of an elastic net. This approach is inspired by the naïve elastic criterion method proposed by (Zou and Hastie, 2005) for the estimation of regression coefficients, from the frequentist viewpoint and the graphical elastic net in (Kovács et al., 2021). Here, the aim is to extend the naïve elastic concept to the Bayesian Gaussian graphical modelling realm, under absolutely continuous priors.

By incorporating both ℓ_1 and ℓ_2 constraints from (2.1) and (3.1) respectively, a Bayesian naïve graphical elastic net is formulated. The hierarchical form is given by

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$
\begin{aligned}
p\left(\mathbf{y}_i\,|\boldsymbol{\Omega}\right) &= N_p\left(\mathbf{y}_i\,|\mathbf{0},\,\boldsymbol{\Omega}^{-1}\right)\quad (i=1,\cdots,n)\\
p\left(\boldsymbol{\Omega}|\mu=0,\sigma,\lambda\right) =& C^{-1}\prod_{i<j}\left\{\mathrm{DE}(\omega_{ij}|\lambda)\right\}\prod_{i=1}^{p}\left\{\mathrm{EXP}(\omega_{ii}|\lambda)\right\}\\
&\times\prod_{i<j}\left\{\mathrm{N}(\omega_{ij}|\sigma)\right\}\prod_{i=1}^{p}\left\{\mathrm{TN}(\omega_{ii}|\sigma)\right\}\mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^{+}}.
\end{aligned}
\tag{4.1}
$$

The normalising constant C^{-1} in this form is dependent on σ and λ, however, techniques similar to those used in (2.20) and (3.10) can be useful here. The posterior is given by

$$
\begin{aligned}
p(\boldsymbol{\Omega}|\mathbf{Y},\mu=0,\sigma,\lambda) \propto& \prod_{i=1}^{n}\det\boldsymbol{\Omega}^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathbf{y}_i^T\boldsymbol{\Omega}\mathbf{y}_i\right\}\\
&\times\prod_{i<j}\left\{\frac{\lambda}{2}\exp(-\lambda|\omega_{ij}|)\right\}\prod_{i=1}^{p}\left\{\frac{\lambda}{2}\exp(-\frac{\lambda}{2}\omega_{ii})\right\}\\
&\times\prod_{i<j}\left\{\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\omega_{ij}^2}{2\sigma^2}\right)\right\}\prod_{i=1}^{p}\left\{\frac{\sqrt{2}}{\sqrt{\pi}\sigma}\exp\left(-\frac{\omega_{ii}^2}{2\sigma^2}\right)\right\}\mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^{+}}.
\end{aligned}
\tag{4.2}
$$

The log of the posterior is given by

$$
\begin{aligned}
p_\ell(\boldsymbol{\Omega}|\mathbf{Y},\mu=0,\sigma,\lambda) &\propto \frac{n}{2}\log(\det\boldsymbol{\Omega})-\frac{1}{2}\mathrm{trace}\left(\mathbf{S}\boldsymbol{\Omega}\right)-\sum_{i=1}^{p}\frac{\lambda}{2}\omega_{ii}-\sum_{i\neq j}\frac{\lambda}{2}|\omega_{ij}|-\frac{1}{2\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{p}\omega_{ij}^2\\
&\propto \log\left(\det\boldsymbol{\Omega}\right)-\mathrm{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right)-\sum_{i<j}\frac{\lambda}{n}|\omega_{ij}|-\sum_{i=1}^{p}\frac{\lambda}{n}\omega_{ii}-\frac{1}{n\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{p}\omega_{ij}^2,
\end{aligned}
\tag{4.3}
$$

where $\mathbf{S}=\mathbf{Y}'\mathbf{Y}$ denotes the sample covariance matrix. The value of $\boldsymbol{\Omega}$ which maximises (4.3) is given by

$$
\underset{\boldsymbol{\Omega}\in\mathbb{M}^{+}}{\arg\max}\left\{\log(\det\boldsymbol{\Omega})-\mathrm{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right)-\rho_1\|\boldsymbol{\Omega}\|_1-\rho_2\|\boldsymbol{\Omega}\|_2\right\},
\tag{4.4}
$$

where $\rho_1=\lambda/n$ and $\rho_2=n/\sigma^2$. Finally, as in Subsection 2.1.1 of Chapter 2, the double exponential distribution in (4.1) can be represented as a scale mixture of Gaussians leading to the hierarchical representation

$$
\begin{aligned}
p(\boldsymbol{\Omega}\,|\boldsymbol{\tau},\mu=0,\sigma,\lambda) &= C_{\boldsymbol{\tau}^*}^{-1}\prod_{i<j}\left\{\frac{1}{\sqrt{2\pi\tau_{ij}}}\exp(-\frac{\omega_{ij}^2}{2\tau_{ij}})\right\}\prod_{i=1}^{p}\left\{\frac{\lambda}{2}\exp(-\frac{\lambda}{2}\omega_{ii})\right\}\\
&\times\prod_{i<j}\left\{\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\omega_{ij}^2}{2\sigma^2}\right)\right\}\prod_{i=1}^{p}\left\{\frac{\sqrt{2}}{\sqrt{\pi}\sigma}\exp\left(-\frac{\omega_{ii}^2}{2\sigma^2}\right)\right\}\mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^{+}}\\
p(\boldsymbol{\tau}|\mu=0,\sigma,\lambda) &\propto C_{\boldsymbol{\tau}^*}\prod_{i<j}\frac{\lambda^2}{2}\exp(-\frac{\lambda^2}{2}\tau_{ij}),
\end{aligned}
\tag{4.5}
$$

where $\boldsymbol{\tau}=\{\tau_{ij}\}_{i<j}$ is a vector of the scale parameters. The intractable normalising constant, $C_{\boldsymbol{\tau}^*}$, depends on $\boldsymbol{\tau}$, λ and σ and has no closed-form solution. The mixing density for τ_{ij} is proportional to an exponential density with rate parameter $\lambda^2/2$.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

4.2 Developing a block Gibbs sampler

The block Gibbs sampler described in Sections 2.3 and 3.2 in Chapters 2 and 3, respectively, provide the foundational building blocks for the sampler described here. The target (posterior) distribution associated with the hierarchical representation in (4.5) can be expressed as

$$\begin{aligned} p(\boldsymbol{\Omega},\boldsymbol{\tau}|\mathbf{Y},\mu=0,\sigma,\lambda) \propto \det\boldsymbol{\Omega}^{\frac{n}{2}} \exp\{-\text{trace}(\frac{1}{2}\mathbf{S}\boldsymbol{\Omega})\} \\ \times\prod_{i<j}\left\{\tau_{ij}^{-\frac{1}{2}}\exp(-\frac{\omega_{ij}^2}{2\tau_{ij}})\exp(-\frac{\lambda^2}{2}\tau_{ij})\right\}\prod_{i=1}^{p}\left\{\exp(-\frac{\lambda}{2}\omega_{ii})\right\} \\ \times\prod_{i<j}\left\{\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\omega_{ij}^2}{2\sigma^2}\right)\right\}\prod_{i=1}^{p}\left\{\frac{\sqrt{2}}{\sqrt{\pi}\sigma}\exp\left(-\frac{\omega_{ii}^2}{2\sigma^2}\right)\right\}\mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^+}. \end{aligned} \tag{4.6}$$

Analogous to the block Gibbs samplers described in Sections 2.3 and 3.2 in Chapters 2 and 3, respectively, the sampler here enables posterior computation of $\boldsymbol{\Omega}$ from the target distribution in (4.6). The Gibbs sampler sequentially updates $\boldsymbol{\Omega}$ one column at time. Without loss of generality, consider updating the last column and row of $\boldsymbol{\Omega}$ and let $\boldsymbol{\Upsilon}^{EN}$ be a symmetric $p\times p$ matrix with a zero main diagonal and $\boldsymbol{\tau}^{EN}$ in the upper and lower off-diagonal entries. Here, $\boldsymbol{\tau}^{EN}$ represents a vector of functions of τ_{ij} and σ^2 values - details are provided below. The exponent 'EN' is used purely for styling in order distinguish from the notation used by the Bayesian graphical lasso and ridge Gibbs samplers. Partition the matrices $\boldsymbol{\Omega}$, $\mathbf{S}$ and $\boldsymbol{\Upsilon}^{EN}$ as follows

$$\underset{p\times p}{\boldsymbol{\Omega}} = \begin{pmatrix} \underset{(p-1)\times(p-1)}{\boldsymbol{\Omega}_{11}} & \underset{(p-1)\times 1}{\boldsymbol{\omega}_{12}} \\ \underset{1\times(p-1)}{\boldsymbol{\omega}_{21}} & \underset{1\times 1}{\omega_{22}} \end{pmatrix}, \quad \underset{p\times p}{\mathbf{S}} = \begin{pmatrix} \underset{(p-1)\times(p-1)}{\mathbf{S}_{11}} & \underset{(p-1)\times 1}{\boldsymbol{s}_{12}} \\ \underset{1\times(p-1)}{\boldsymbol{s}_{21}} & \underset{1\times 1}{s_{22}} \end{pmatrix}, \quad \underset{p\times p}{\boldsymbol{\Upsilon}^{EN}} = \begin{pmatrix} \underset{(p-1)\times(p-1)}{\boldsymbol{\Upsilon}_{11}^{EN}} & \underset{(p-1)\times 1}{\boldsymbol{\tau}_{12}^{EN}} \\ \underset{1\times(p-1)}{\boldsymbol{\tau}_{21}^{EN}} & \underset{1\times 1}{0} \end{pmatrix}. \tag{4.7}$$

Recall that the $\det\boldsymbol{\Omega}^{\frac{n}{2}}$ can be represented as

$$\det\boldsymbol{\Omega}^{\frac{n}{2}} = (\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}}\det\boldsymbol{\Omega}_{11}^{\frac{n}{2}} \propto (\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}},$$

since we are only interested in the last column and row. Similarly,

$$\text{trace}\left(\frac{1}{2}\mathbf{S}\boldsymbol{\Omega}\right) \propto -\frac{1}{2}(2\boldsymbol{s}_{21}\boldsymbol{\omega}_{12}+s_{22}\omega_{22}).$$

Notice that the conditional distribution of the last column in (4.6) can be presented as

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$
\begin{aligned}
p(\omega_{12},\omega_{22}\,|\,\boldsymbol{\Omega}_{11},\boldsymbol{\Upsilon}^{EN},\mathbf{S},\mu,\sigma,\lambda) &\propto (\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}}\exp\left(\frac{1}{2}(2\mathbf{s}_{21}\boldsymbol{\omega}_{12}+s_{22}\omega_{22})\right)\\
&\times\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{p-1}\frac{\omega_{ij}^2}{\tau_{ij}}+\sum_{i=1}^{p-1}\frac{\omega_{ij}^2}{\sigma_{ij}^2}+\omega_{22}\lambda+\left(\frac{\omega_{22}}{\sigma_{22}}\right)^2\right]\right\}\\
&=(\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}}\exp(\frac{1}{2}(2\boldsymbol{s_{21}}\boldsymbol{\omega}_{12}+s_{22}\omega_{22}))\\
&\times\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{p-1}\left(\frac{\tau_{ij}+\sigma_{ij}^2}{\tau_{ij}\sigma_{ij}^2}\right)\omega_{ij}+\omega_{22}\lambda+\left(\frac{\omega_{22}}{\sigma_{22}}\right)^2\right]\right\}\\
&=(\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12})^{\frac{n}{2}}\\
&\times\exp\left[-\frac{1}{2}\left\{\boldsymbol{\omega}_{21}\boldsymbol{D}^{EN}\boldsymbol{\omega}_{12}+2\boldsymbol{s}_{21}\boldsymbol{\omega}_{12}+\omega_{22}\left(\lambda+s_{22}+\frac{\omega_{22}}{\sigma_{22}^2}\right)\right\}\right]\\
&=\left(\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}\right)^{\frac{n}{2}}\\
&\times\exp\left[-\frac{1}{2}\left\{\boldsymbol{\omega}_{21}\boldsymbol{D}^{EN}\boldsymbol{\omega}_{12}+2\boldsymbol{s}_{21}\boldsymbol{\omega}_{12}+\frac{\omega_{22}}{\sigma^2}\left(\sum_{k=0}^{1}\binom{1}{k}([\lambda+s_{22}]\sigma^2)^{1-k}+\omega_{22}^k\right)\right\}\right]\\
&\propto\left(\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}\right)^{\frac{n}{2}}\\
&\times\exp\left[-\frac{1}{2}\left\{\boldsymbol{\omega}_{21}\boldsymbol{D}^{EN}\boldsymbol{\omega}_{12}+2\boldsymbol{s}_{21}\boldsymbol{\omega}_{12}+\omega_{22}\left(\lambda+s_{22}+\frac{1}{\sigma^2}\right)\right\}\right];\ \text{for } k=0,
\end{aligned}
\tag{4.8}
$$

where $\boldsymbol{D}^{EN}=\text{diag}\{(\sigma^2+\tau)/(\sigma^2\tau)\}$. Consider the following change of variables

$$
\begin{aligned}
\boldsymbol{\beta}&=\boldsymbol{\omega}_{12}\\
\gamma&=\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12},
\end{aligned}
$$

with the constant Jacobian not involving $\boldsymbol{\beta}$ nor γ (refer to Section 2.3 in Chapter 2 for more details). This transformation yields the following conditional distribution

$$
\begin{aligned}
p(\gamma,\boldsymbol{\beta}\,|\,\boldsymbol{\Omega}_{11},\boldsymbol{\Upsilon},\boldsymbol{S},\mu,\sigma,\lambda)&\propto(\gamma)^{\frac{n}{2}}\exp(-\frac{s_{22}+\lambda+1/\sigma^2}{2}\gamma)\\
&\times\exp\left[-\frac{1}{2}(\boldsymbol{\beta}^\top\{\boldsymbol{\tau}^{-1}+(s_{22}+\lambda+1/\sigma^2)\boldsymbol{\Omega}_{11}^{-1}\}\boldsymbol{\beta}+2\boldsymbol{s}_{21}\boldsymbol{\beta})\right].
\end{aligned}
$$

It follows that

$$
(\gamma,\boldsymbol{\beta}|\,\boldsymbol{\Omega}_{11},\boldsymbol{\Upsilon},\boldsymbol{S},\lambda)\sim\text{GA}(\frac{n}{2}+1,\frac{s_{22}+\lambda+1/\sigma^2}{2})\text{N}_{p-1}(-\boldsymbol{C}\boldsymbol{s}_{21},\boldsymbol{C}),
$$

where $\boldsymbol{C}=\{(s_{22}+\lambda+1/\sigma^2)\boldsymbol{\Omega}_{11}+\boldsymbol{D}^{EN}\}^{-1}$, GA$(\cdot)$ refers to a gamma pdf and N$_{p-1}(\cdot)$ to a $(p-1)$-variate Gaussian pdf. Updating $\boldsymbol{\tau}$ is achieved by using methodologies described in Section 2.3 in Chapter 2. In particular, the conditional posterior distributions for the $\psi_{ij}=1/\tau_{ij}$ is given by the inverse Gaussian pdf

$$
p(x;\mu^*,\lambda^*)=\sqrt{\frac{\lambda^*}{2\pi}}x^{-\frac{3}{2}}\exp\left(-\frac{\lambda^*(x-\mu^*)^2}{2(\mu^*)^2x}\right)\ x>0,
$$

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

where $\mu* = \sqrt{\lambda^2/\omega_{ij}^2}$ and $\lambda^* = \lambda^2$. The block Gibbs sampler is summarised in Algorithm 5. It is also worth mentioning that this sampler upholds the positive definiteness restriction on $\boldsymbol{\Omega}$, again refer to 2.3 in Chapter 2 for a detailed description.

Algorithm 5 Naïve Bayesian graphical elastic net: Block Gibbs sampler

Require: Initialise $\boldsymbol{\Omega}$.
  for $i = 1, \ldots, p$ **do**
    1) Partition $\boldsymbol{\Omega}$, $\mathbf{S}$ and $\boldsymbol{\Upsilon}^{EN}$ as in (4.7).
    2) Sample $\gamma \sim \text{GA}(n/2+1, (s_{22}+\lambda+1/\sigma^2)/2)$ and $\boldsymbol{\beta} \sim \text{N}_{\text{p}-1}(-\boldsymbol{C}\boldsymbol{s}_{21}, \boldsymbol{C})$.
    3) Update $\boldsymbol{\omega}_{12} = \boldsymbol{\beta}$, $\boldsymbol{\omega}_{21} = \boldsymbol{\beta}^\top$ and $\omega_{22} = \gamma + \boldsymbol{\beta}^\top \boldsymbol{\Omega_{11}}^{-1} \boldsymbol{\beta}$.
  end for
  for $i \neq j$ **do**
    Sample $\psi_{ij} \sim \text{IG}(\mu^*, \lambda^*)$ where $\lambda^* = \lambda^2$ and $\mu^* = \sqrt{(\lambda^2/\omega_{ij}^2)}$ and update $\tau_{ij} = 1/\psi_{ij}$.
  end for

4.2.1 Computational insights

Given the archetypal similarities between the block Gibbs sampler developed here and in that proposed in Chapter 3, the convergence and inefficiency characteristics are near identical. For the sake of brevity these results have been omitted here.

4.3 Naïve Bayesian graphical adaptive elastic net

Similar to the Bayesian graphical lasso and Bayesian graphical ridge-type priors in (2.2) and (3.2), respectively, the naïve Bayesian graphical elastic requires the selection of σ and λ. The block Gibbs sampler in Section 4.2, as well as the hierarchical forms in (4.1) and (4.5) allows for extended hierarchical formulations of the prior to allow for adaptive shrinkage on different off-diagonal elements of $\boldsymbol{\Omega}$. That is, the Bayesian framework facilitates the usage of a hyperprior on σ and λ, and extending the block Gibbs samplers to sample from the posterior distribution thereof. Within this setting, however, the set of priors for the naïve Bayesian adaptive graphical elastic net may be derived under two formulations. The first approach subsumes the archetypal formulations in (2.20) and (3.10). In particular,

$$
\begin{aligned}
p\left(\mathbf{y}_i \,|\boldsymbol{\Omega}\right) &= N_p\left(\mathbf{y}_i \,|\mathbf{0}, \boldsymbol{\Omega}^{-1}\right) \quad (i = 1, \cdots, n) \\
p\left(\boldsymbol{\Omega} \,|\mu = 0, \{\kappa_{ij}, \lambda_{ij}\}_{i<j}\right) &= C^{-1}_{\{\kappa_{ij}, \lambda_{ij}\}_{i<j}} \prod_{i<j} \left\{\text{DE}(\omega_{ij} \,|\, \lambda_{ij})\right\} \prod_{i=1}^{p} \left\{\text{EXP}(\omega_{ii} \,|\, \lambda_{ii})\right\} \\
&\quad \times \prod_{i<j} \left\{\text{N}(\omega_{ij} \,|\kappa_{ij}^{\frac{1}{2}})\right\} \prod_{i=1}^{p} \left\{\text{TN}(\omega_{ii} \,|\kappa_{ii}^{\frac{1}{2}})\right\} \mathbb{1}_{\boldsymbol{\Omega} \in \mathbb{M}^+} \qquad (4.9) \\
p\left(\{\kappa_{ij}\}_{i<j} \,|\{\kappa_{ii}\}_{i=1}^{p}, \lambda\right) &\propto C_{\{\kappa_{ij}\}_{i<j}} \prod_{i<j} \text{IGA}(a, b) \\
p(\{\lambda_{ij}\}_{i<j} \,|\, \{\lambda_{ii}\}_{i=1}^{p}, \kappa) &\propto C_{\{\lambda_{ij}\}_{i<j}} \prod_{i<j} \text{GA}(r, s).
\end{aligned}
$$

Here, $\kappa = \sigma^2$. The set of priors in (4.9) make use of different shrinkage parameters λ_{ij} and κ_{ij} on each off-diagonal element ω_{ij}. Secondly, $\{\lambda_{ii}, \kappa_{ii}\}_{i=1}^{p}$ for the diagonal elements are treated as hyperparameters. Finally, $C_{\{\lambda_{ij}\}_{i\leq j}}$ and

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$C_{\{\kappa_{ij}\}_{i\leq j}}$ are the intractable normalisation constants dependent on λ_{ij} and κ_{ij}, and $C^{-1}_{\{\kappa_{ij},\lambda_{ij}\}_{i<j}} = C^{-1}_{\{\kappa_{ij}\}_{i<j}} C^{-1}_{\{\lambda_{ij}\}_{i<j}}$. The posterior of the $\{\lambda_{ij}\}_{i<j}$ and $\{\kappa_{ij}\}_{i<j}$ conditioned on $\boldsymbol{\Omega}$ is given by (2.21) and (3.11), respectively. An obvious caveat here is the presence of four hyperparameters. Selecting these hyperparameters can introduce computational complexity, impacting its parsimonious nature, however, it may enable finer tuning as well. An alternative formulation in favour of parsimony is given by

$$
\begin{aligned}
p\left(\mathbf{y}_i \,|\boldsymbol{\Omega}\right) &= N_p\left(\mathbf{y}_i \,|\mathbf{0}, \boldsymbol{\Omega}^{-1}\right) \quad (i = 1, \cdots, n) \\
p\left(\boldsymbol{\Omega} \,|\mu = 0, \{\eta_{ij}, \lambda_{ij}\}_{i<j}\right) &= C^{-1}_{\{\eta_{ij},\lambda_{ij}\}_{i<j}} \prod_{i<j}\left\{\mathrm{DE}(\omega_{ij} \,|\, \lambda_{ij})\right\} \prod_{i=1}^{p}\left\{\mathrm{EXP}(\omega_{ii} \,|\, \lambda_{ii})\right\} \\
&\quad \times \prod_{i<j}\left\{\mathrm{N}(\omega_{ij} \,|\eta_{ij}^{-\frac{1}{2}})\right\} \prod_{i=1}^{p}\left\{\mathrm{TN}(\omega_{ii} \,|\eta_{ii}^{-\frac{1}{2}})\right\} \mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^+} \qquad (4.10) \\
p\left(\{\eta_{ij}\}_{i<j} \,|\{\eta_{ii}\}_{i=1}^{p}, \lambda\right) &\propto C_{\{\eta_{ij}\}_{i<j}} \prod_{i<j} \mathrm{EXP}(b) \\
p(\{\lambda_{ij}\}_{i<j} \,|\, \{\lambda_{ii}\}_{i=1}^{p}, \eta) &\propto C_{\{\lambda_{ij}\}_{i<j}} \prod_{i<j} \mathrm{EXP}(s).
\end{aligned}
$$

Here, $\eta = \frac{1}{\sigma^2}$ and the intractable normalisation constants simplify as before. This choice of prior formulation aims to benefit from the strengths of the lasso and ridge components, while avoiding an increase in the number of necessary hyperparameters. Conditional on $\boldsymbol{\Omega}$, the target distributions for λ_{ij} and η_{ij} are given by

$$
\begin{aligned}
p(\lambda_{ij} \,|\{\lambda_{ii}\}_{i=1}^{p}, \boldsymbol{\Omega}, \eta) &= \frac{\lambda_{ij}}{2} \exp(-\lambda_{ij}|\omega_{ij}|) b \exp(-b\lambda_{ij}) \\
&\propto \frac{\left(|\omega_{ij}| + b\right)^2}{\Gamma(2)} \lambda_{ij} \exp\left\{-\lambda_{ij}(|\omega_{ij}| + b)\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
p\left(\eta_{ij} \,|\{\eta_{ii}\}_{i=1}^{p}, \boldsymbol{\Omega}, \lambda\right) &= \sqrt{\frac{\eta_{ij}}{2\pi}} \exp\left(-\frac{\omega_{ij}^2 \eta_{ij}}{2}\right) s \exp(-s\eta_{ij}) \\
&\propto \frac{\left(\frac{\omega_{ij}^2}{2} + s\right)^{\frac{3}{2}}}{\Gamma(\frac{3}{2})} \eta_{ij}^{\frac{1}{2}} \exp\left\{-\eta_{ij}\left(\frac{\omega_{ij}^2}{2} + b\right)\right\}.
\end{aligned}
$$

In other words,

$$
\lambda_{ij} \,|\left(\{\lambda_{ii}\}_{i=1}^{p}, \boldsymbol{\Omega}, \eta\right) \sim \mathrm{GA}(2, |\omega_{ij}| + s) \qquad (4.11)
$$

$$
\eta_{ij} \,|\left(\{\eta_{ii}\}_{i=1}^{p}, \boldsymbol{\Omega}, \lambda\right) \sim \mathrm{GA}\left(\frac{3}{2}, \frac{\omega_{ij}^2}{2} + b\right). \qquad (4.12)
$$

The hyperparameters in (4.9), and in (4.10) are selected using a grid search approach analogous to that described in Section 3.4 of Chapter 3. For each permutation considered, the hyperparameters are determined by the mean values

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

of the lowest 5% of Stein's loss values across all topologies and for $p \in \{10, 30, 50, 100\}$ and $n \in \{5, \cdots, 200\}$.

Algorithm 6 Type I naïve Bayesian adaptive graphical elastic net: Block Gibbs sampler

Require: Initialise $\boldsymbol{\Omega}$.
for $i = 1, \ldots, p$ **do**
1) Partition $\boldsymbol{\Omega}$, $\mathbf{S}$ and $\boldsymbol{\Upsilon}^{EN}$ as in (4.7).
2) Sample $\gamma \sim \text{GA}(n/2+1, (s_{22}+\lambda_{ij}+1/\kappa_{ij})/2)$ and $\boldsymbol{\beta} \sim \text{N}_{p-1}(-\boldsymbol{C}\boldsymbol{s}_{21}, \boldsymbol{C})$.
3) Update $\boldsymbol{\omega}_{12} = \boldsymbol{\beta}$, $\boldsymbol{\omega}_{21} = \boldsymbol{\beta}^\top$ and $\omega_{22} = \gamma + \boldsymbol{\beta}^\top \boldsymbol{\Omega_{11}}^{-1} \boldsymbol{\beta}$.
end for
for $i \neq j$ **do**
1) Sample $\lambda_{ij} \sim \text{GA}(r+1, |\omega_{ij}|+s)$.
2) Sample $\kappa_{ij} \sim \text{IGA}(a+\frac{1}{2}, b+\frac{\omega_{ij}^2}{2})$.
3) Sample $\psi_{ij} \sim \text{IG}(\mu^*, \lambda^*)$ where $\lambda^* = \lambda_{ij}^2$ and $\mu^* = \sqrt{(\lambda_{ij}^2/\omega_{ij}^2)}$ and update $\tau_{ij} = 1/\psi_{ij}$.
end for

Algorithm 7 Type II naïve Bayesian adaptive graphical elastic net: Block Gibbs sampler

Require: Initialise $\boldsymbol{\Omega}$.
for $i = 1, \ldots, p$ **do**
1) Partition $\boldsymbol{\Omega}$, $\mathbf{S}$ and $\boldsymbol{\Upsilon}^{EN}$ as in (4.7).
2) Sample $\gamma \sim \text{GA}(n/2+1, (s_{22}+\lambda_{ij}+\eta_{ij})/2)$ and $\boldsymbol{\beta} \sim \text{N}_{p-1}(-\boldsymbol{C}\boldsymbol{s}_{21}, \boldsymbol{C})$.
3) Update $\boldsymbol{\omega}_{12} = \boldsymbol{\beta}$, $\boldsymbol{\omega}_{21} = \boldsymbol{\beta}^\top$ and $\omega_{22} = \gamma + \boldsymbol{\beta}^\top \boldsymbol{\Omega_{11}}^{-1} \boldsymbol{\beta}$.
end for
for $i \neq j$ **do**
1) Sample $\lambda_{ij} \sim \text{GA}(2, |\omega_{ij}|+s)$.
2) Sample $\eta_{ij} \sim \text{GA}\left(\frac{3}{2}, \frac{\omega_{ij}^2}{2}+b\right)$.
3) Sample $\psi_{ij} \sim \text{IG}(\mu^*, \lambda^*)$ where $\lambda^* = \lambda_{ij}^2$ and $\mu^* = \sqrt{(\lambda_{ij}^2/\omega_{ij}^2)}$ and update $\tau_{ij} = 1/\psi_{ij}$.
end for

4.4 Synthetic data analysis

The synthetic study from Section 3.5 in Chapter 3 is revisited here. Recall that this experimental setup aims to evaluate the performance of the precision estimates across the varying network topologies in Section 3.5 in Chapter 3. The p/n ratios considered here are $p/n \in \{0.5, 1, 4, 10\}$ for $p = 100$. This configuration allows for extensive testing in scenarios where $n > p$, $n = p$ and $n < p$, covering both low and high-dimensional network modelling. The Bayesian estimators considered include the Bayesian adaptive graphical lasso in (2.20), the Bayesian adaptive graphical ridge-type in (3.10), the Bayesian graphical horseshoe from (Li et al., 2019) and the naïve Bayesian adaptive graphical elastic net in (4.9) and (4.10). The frequentist graphical elastic net estimator from (Kovács et al., 2021) is also included. The estimators are denoted as BAGL, BAGR, BGHS, BAGENI, BAGENII and GELNET respectively. To assist the reading experience, Table 4.1 provides a summary of the algorithms considered in this study.

Recall from Section 2.2 in Chapter 2, and Section 3.6 in Chapter 3 that the continuous priors of the Bayesian estimators (BAGL, BAGR, BAGENI, BAGENII , and BGHS in this case) require a heuristic treatment for edge selection since these methods are not capable of estimating elements of $\boldsymbol{\Omega}$ to be exactly equal to zero. Two popular techniques include a thresholding approach where an edge is determined based if the partial correlation or function thereof is greater than some pre-defined bound. A Wishart treatment is introduced by (Wang, 2012), described in Chapter 2 and briefly explored in Chapter 3. The second approach is centered around the use of credible intervals obtained from the posterior distribution. This technique has enjoyed success in ((Khondker et al., 2013); (Li et al., 2019) and (Sagar et al., 2021)), and is the technique of choice in this simulation study. This approach constructs an $(1-\delta)\%$, where $0 \leq \delta \leq 1$, symmetric interval for each $\hat{\omega}_{ij}$ for some user defined δ, and an edge is excluded if the interval includes zero. This study considers $(1-\delta)\% = 90\%$.

Model configuration. The BAGL, BAGR, BAGENI and BAGENII estimators are implemented via the 'baygel' R package. The parameters of the gamma priors in the BAGL estimator are set to $r = 10^{-2}$ and $s = 10^{-6}$, as proposed by the founding author. Additionally, the parameters of the inverse gamma priors for the BAGR estimator

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 4.1 *Summary of the precision matrix algorithms considered in the synthetic data analysis study.*

Abbreviation	Full name	Author
BAGL	Bayesian adaptive graphical lasso	(Wang, 2012)
BAGR	Bayesian adaptive graphical ridge	(Smith et al., 2022a)
BAGENI	Type I naïve Bayesian adaptive graphical elastic net	(Smith et al., 2023)
BAGENII	Type II naïve Bayesian adaptive graphical elastic net	(Smith et al., 2023)
BGHS	Bayesian graphical horseshoe	(Li et al., 2019)
GELNET	Graphical elastic net	(Kovács et al., 2021)

are set to $a = 1$ and $b = 10^{-2}$ as defined in Section 3.4 in Chapter 3. The parameters of the inverse gamma and gamma priors for the BAGENI estimator are set to $r = 10^{-3}$, $s = 10^{-2}$, $a = 10^{-3}$ and $b = 10^{-1}$. The parameters of the exponential priors for the BAGENII estimator are set to $s = 10^{-1}$ and $b = 10^{-3}$. The BGHS estimator is implemented via the 'GHS' R package. Finally, the GELNET estimator is implemented via the 'GLassoElnetFast' R package. The penalty parameter of the GELNET estimator is determined using five-fold cross validation on each topology considered, and the tuning parameter is set to $\alpha = 0.5$ which equally weighs the ℓ_1 and ℓ_2 penalty terms in estimator's objective function. Lastly, the identity matrix is used as the target for the GELNET estimator given its balanced performance within the simulation studies performed by the founding authors.

Loss functions. The loss functions used within this simulation are given by

- Stein's loss: $\text{trace}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Omega}) - \log(|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Omega}|) - p$.
- Frobenius norm: $\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\|_F$, where $\|\boldsymbol{A}\|_F = \sqrt{\text{trace}(\boldsymbol{A}\boldsymbol{A}^\top)}$ and referred to as F norm within the study.

Performance functions. The network recovery for all estimates are assessed based on

- True positive (TP): an edge is correctly predicted.
- False positive (FP): an edge is incorrectly predicted.
- True negative (TN): the absence of an edge is correctly predicted.
- False negative (FN): the absence of an edge is incorrectly predicted.
- Sensitivity: $\frac{\text{TP}}{\text{TP+FN}}$.
- Specificity: $\frac{\text{TN}}{\text{TN+FP}}$.
- Matthews correlation coefficient (MCC): $\frac{\text{TP}\times\text{TN}-\text{FP}\times\text{FN}}{\sqrt{(\text{TP+FP})(\text{TP+FN})(\text{TN+FP})(\text{TN+FN})}}$.

Higher values for sensitivity, specificity and MCC indicate superior performance. A description on the scope of these metrics is provided in Section 3.5 of Chapter 3.

Simulation results. The results presented in Tables 4.2 - 4.5 offer valuable insights into the behaviour of the estimation techniques considered. Before perusing the results a note on styling: the results tables include performance metrics which assess, for all estimators, both signal strength (value of $\hat{\omega}_{ij}$ recovery and network structure recovery. The standard errors for the network recovery metrics across all estimators are below 0.05, with the exception of the BAGL estimator which saw standard errors up to 0.1. For this reason the standard errors for these metrics are omitted. Next, across all metrics, the best performing metric is boldfaced, however, this is only applicable when two or less estimators obtain the podium position. Finally, given that the BAGENI and BAGENII

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

estimators are archetypally composed of both the BAGL and BAGR constituents, the top two performing signal strength recovery metrics in this comparison are highlighted in order to assess the adaptability of novel estimator.

First, considering the setting where $n > p$ in Table 4.2, the BGHS enjoys the lowest loss values for sparse topolgies and the BAGR for the denser ones. A similar trend is observed where $n = p$ in Table 4.3.

Second, where $n < p$ in Table 4.4, there is a mixed bag of results. The BGHS is the top performer in the AR(1) and circle (under Stein) topology. The GELNET is the top performer in the scale-free, random (under Stein) and cluster (under Stein) topologies. The BAGENII is the top performer in the star, circle (under F norm), random (under F norm), full (under F norm) topologies. The BAGR is the top performer in the AR(2), random (under F norm), cluster (under F norm), hub (under F norm), band (under F norm) and full (under Stein) topologies. The BAGL struggles in this setting.

Third, where $n << p$ in Table 4.5, the BAGR is generally the top performer with the BAGENII closely behind. The BAGL and BGHS estimators struggle in this scenario but the GELNET displays commendable performance.

Fourth, the BAGENI, BAGENII, BAGR and GELNET estimators enjoy consistent low standard errors accros all topologies and p/n ratios. The latter showing slight signs of struggle for $p/n = 10$.

Fifth, the BAGENII estimator generally outperforms the BAGENI estimator, especially when $n << p$. This suggests that the parsimonious estimator generalises better than its counterpart under the testing environments considered here.

In terms of variable selection, the results showcase similar behaviour to their numerical loss counterparts. Considering the setting where $n \geq p$, the BGHS generally has the strongest performance, particulalrly when $n > p$, followed closely by the BAGENII and BAGL estimators. The denser topologies prove to be problematic for all estimators but this could be due to the small magnitude of the signal size in the true Ω. The GELNET appears to struggle the most, according to the MCC results, across all toplogies. When $n < p$, the BAGENII, BAGR and BGHS are the top performing estimators with the BAGENI closely behind. For $n << p$, the BAGR and both BAGENI and BAGENII estimators are generally the top performing estimators, whilst the BAGL, BGHS and GELNET appear to struggle under these conditions. Reviewing the results holistically, the specificity of the dense toplogies are unsurprisingly low due the absence of of true negatives - this agrees with the insights derived from the simulation study results in Chapter 3.

Lastly, when considering the BAGL, BAGR, BAGENI and BAGENII estimators, the BAGENII estimator illustrates an impressive adaptability between the former two. In particular, BAGENII is consistently either the best or runner up performer in this comparison. These results illustrates its ability to adapt to the inherent sparsity level in the topologies considered and to the scarcity of available data and is therefore the recommended estimator when apriori knowledge about these conditions are unknown. The BAGENII estimator also displays commendable performance when compared to the start-of-the-art BGHS and GELNET estimators.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 4.2 *Summary of network signal strength and structure recovery performance across multiple network topologies and estimators for* $p/n = 0.5$. The median performance values and (standard errors) reported here are based on 30 replications for the BAGENI, BAGENII, BAGL, BAGR, BGHS and GELNET estimators. The best performance value is boldfaced. Cell colours correspond only to the BAGL, BAGR, BAGENI and BAGENII signal strength metrics, with darker colours indicating the best performance value and lighter colours representing the runner-up.

$\frac{p}{n} = 0.5$	**AR(1)**						**AR(2)**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	16.5	18.7	**10.8**	22.2	14.4	46.6	11.4	8.00	4.61	13.6	**3.85**	16.8
	(0.37)	(0.41)	(0.53)	(0.36)	(0.63)	(0.80)	(0.39)	(0.28)	(0.25)	(0.34)	(0.25)	(0.33)
F norm	**7.75**	10.1	11.69	9.58	10.3	12.1	4.54	3.66	2.54	4.86	**2.46**	5.61
	(0.13)	(0.08)	(0.23)	(0.06)	(0.13)	(0.06)	(0.12)	(0.10)	(0.10)	(0.07)	(0.12)	(0.05)
Specificity	**0.46**	0.45	0.45	0.45	0.45	0.44	1	1	1	1	1	0.99
Sensitivity	0.89	0.93	0.85	0.81	**1**	0.99	0.72	0.91	0.91	0.59	**0.98**	0.52
MCC	0.19	0.18	**0.2**	0.15	0.19	0.15	0.83	0.88	0.93	0.72	**0.95**	0.63
$\frac{p}{n} = 0.5$	**star**						**circle**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	9.23	3.08	4.55	7.62	**2.55**	18.6	11.8	7.89	3.85	13.4	**2.17**	49.3
	(0.32)	(0.15)	(0.30)	(0.19)	(0.26)	(2.54)	(0.30)	(0.21)	(0.21)	(0.28)	(0.17)	(0.37)
F norm	6.12	2.67	3.59	4.65	**2.38**	2.07	4.43	2.28	1.84	3.99	**1.30**	5.89
	(0.17)	(0.07)	(0.16)	(0.06)	(0.15)	(0.14)	(0.17)	(0.11)	(0.18)	(0.07)	(0.12)	(0.03)
Specificity	0.99	0.99	0.99	0.99	0.99	**1**	1	1	1	1	1	1
Sensitivity	0.46	0.88	0.7	0.43	**0.93**	0.06	0.56	0.88	0.94	0.35	**1**	0.12
MCC	0.53	0.82	0.69	0.53	**0.84**	0.18	0.74	0.92	0.97	0.57	**1**	0.3
$\frac{p}{n} = 0.5$	**scale-free**						**random**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	8.91	3.81	4.26	7.91	**2.82**	3.93	8.78	4.81	4.44	8.50	**3.52**	6.17
	(0.27)	(0.18)	(0.24)	(0.15)	(0.21)	(0.23)	(0.30)	(0.33)	(0.36)	(0.25)	(0.36)	(0.46)
F norm	6.27	3.53	3.71	5.18	**2.79**	4.11	6.18	3.76	3.93	5.06	**3.17**	4.75
	(0.15)	(0.22)	(0.18)	(0.17)	(0.16)	(0.18)	(0.17)	(0.17)	(0.19)	(0.13)	(0.17)	(0.25)
Specificity	0.99	0.99	0.99	0.99	0.99	**1**	0.99	0.99	0.99	0.99	0.99	**1**
Sensitivity	0.55	91	0.78	0.53	**97**	0.27	0.7	0.94	0.86	0.7	**0.98**	0.23
MCC	0.64	**0.75**	0.73	0.63	0.74	0.49	0.79	**0.87**	0.86	0.79	**0.87**	0.44
$\frac{p}{n} = 0.5$	**cluster**						**hub**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	25.2	26.4	29.6	**23.9**	33.7	31.2	8.55	3.88	3.81	8.73	**2.29**	6.39
	(0.24)	(0.22)	(0.69)	(0.16)	(0.50)	(0.06)	(0.23)	(0.17)	(0.21)	(0.18)	(0.15)	(0.25)
F norm	28.9	29.1	29.2	**28.7**	29.3	29.2	6.53	4.99	3.86	7.26	**2.89**	7.10
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.13)	(0.15)	(0.14)	(0.05)	(0.12)	(0.08)
Specificity	0.04	0.04	0.04	0.04	0.04	0.04	1	1	1	1	1	1
Sensitivity	0.98	0.99	0.99	0.98	**1**	**1**	0.59	0.91	0.81	0.54	**0.98**	0.18
MCC	0.02	0.02	0.02	0.02	0.02	0.02	0.72	0.88	0.87	0.68	**0.94**	0.39
$\frac{p}{n} = 0.5$	**band**						**full**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	8.06	3.50	3.62	9.04	**2.02**	6.90	46.1	54.0	69.1	**40.4**	82.7	64.7
	(0.23)	(0.18)	(0.18)	(0.21)	(0.18)	(0.44)	(0.35)	(0.32)	(1.15)	(0.34)	(1.27)	(0.01)
F norm	5.99	3.08	3.68	5.31	**2.51**	5.55	49.3	49.6	49.8	**49.0**	49.9	49.7
	(0.14)	(0.09)	(0.18)	(0.06)	(0.16)	(0.19)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Specificity	1	1	1	1	1	1	0	0	0	0	0	0
Sensitivity	0.73	0.97	0.87	0.75	**0.99**	0.14	1	1	1	1	1	1
MCC	0.85	0.98	0.93	0.86	**0.99**	0.33	0	0	0	0	0	0

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 4.3 *Summary of network signal strength and structure recovery performance across multiple network topologies and estimators for* $p/n = 1$. The median performance values and (standard errors) reported here are based on 30 replications for the BAGENI, BAGENII, BAGL, BAGR, BGHS and GELNET estimators. The best performance value is boldfaced. Cell colours correspond only to the BAGL, BAGR, BAGENI and BAGENII signal strength metrics, with darker colours indicating the best performance value and lighter colours representing the runner-up.

$\frac{p}{n} = 1$	**AR(1)**						**AR(2)**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	29.4	30.2	**20.1**	34.9	25.9	63.0	20.6	15.3	10.9	20.2	**10.1**	23.8
	(0.52)	(0.46)	(0.78)	(0.35)	(0.97)	(1.33)	(0.54)	(0.46)	(8.01)	(0.35)	(0.68)	(0.84)
F norm	**9.63**	11.1	10.1	11.3	11.7	13.5	7.03	5.58	4.39	6.16	**4.37**	6.72
	(0.08)	(0.05)	(0.19)	(0.03)	(0.12)	(0.10)	(0.19)	(0.10)	(1.55)	(0.05)	(0.20)	(0.12)
Specificity	0.45	0.45	0.45	0.44	0.44	0.44	0.99	0.99	0.99	0.98	0.99	0.98
Sensitivity	0.85	0.93	0.93	0.82	0.99	**1**	0.72	0.90	0.86	0.71	**0.97**	0.7
MCC	0.15	0.15	**0.16**	0.13	**0.16**	0.13	0.75	0.73	**0.8**	0.69	0.77	0.64
$\frac{p}{n} = 1$	**star**						**circle**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	15.3	**3.69**	11.5	7.73	5.17	12.7	20.8	11.9	105	20.2	**4.32**	49.5
	(0.40)	(0.11)	(1.62)	(0.18)	(0.27)	(2.62)	(0.68)	(0.45)	(5.29)	(0.58)	(0.43)	(0.45)
F norm	9.01	2.76	8.08	4.29	3.71	**2.09**	6.89	2.69	8.79	4.34	**2.13**	5.90
	(0.38)	(0.36)	(0.48)	(0.26)	(0.38)	(0.07)	(0.28)	(0.15)	(1.11)	(0.08)	(0.21)	(0.04)
Specificity	0.98	0.99	0.99	0.98	0.98	**1**	1	1	1	1	1	1
Sensitivity	0.49	**0.95**	0.56	0.72	0.88	0.07	0.63	0.93	0.72	0.52	**1**	0.12
MCC	0.47	**0.71**	0.52	0.59	0.63	0.22	0.78	0.96	0.84	0.71	**1**	0.3
$\frac{p}{n} = 1$	**scale-free**						**random**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	14.5	5.41	8.57	8.81	**5.08**	5.67	14.3	8.30	9.49	11.0	**7.21**	10.0
	(0.40)	(0.11)	(1.62)	(0.18)	(0.27)	(0.59)	(0.44)	(0.53)	(0.56)	(0.40)	(0.57)	(1.05)
F norm	8.97	4.42	6.05	5.39	**4.13**	4.78	8.62	5.12	6.44	5.65	**4.84**	5.74
	(0.23)	(0.24)	(0.28)	(0.25)	(0.17)	(0.23)	(0.22)	(0.23)	(0.33)	(0.17)	(0.24)	(0.38)
Specificity	0.99	0.98	0.99	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.98	**1**
Sensitivity	0.57	**0.96**	0.70	0.77	**0.96**	0.37	0.72	**0.98**	0.79	0.87	0.97	0.3
MCC	0.54	0.61	0.59	0.57	**0.62**	0.5	**0.67**	0.66	**0.67**	0.65	0.64	0.5
$\frac{p}{n} = 1$	**cluster**						**hub**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	31.9	28.9	31.0	**28.2**	33.2	31.3	14.0	6.94	8.53	11.3	**5.61**	9.90
	(0.42)	(0.27)	(0.85)	(0.18)	(0.71)	(0.02)	(0.33)	(0.33)	(7.94)	(0.25)	(0.41)	(0.47)
F norm	29.4	29.2	29.2	**29.0**	29.3	29.3	9.35	6.98	6.66	8.11	**5.17**	7.96
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.18)	(0.14)	(0.37)	(0.05)	(0.26)	(0.10)
Specificity	0.04	0.04	0.04	0.04	0.04	0.04	0.99	0.99	0.99	0.99	0.99	**1**
Sensitivity	0.98	1	0.98	0.99	1	1	0.63	0.95	0.76	0.79	**0.96**	0.21
MCC	0.02	0.02	0.02	0.02	0.02	0.02	0.63	0.7	0.73	0.62	**0.76**	0.43
$\frac{p}{n} = 1$	**band**						**full**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	13.1	7.97	7.88	14.1	**5.08**	12.9	52.1	58.4	66.1	**49.9**	80.1	64.7
	(0.42)	(0.45)	(0.54)	(0.29)	(0.41)	(0.62)	(0.46)	(0.55)	(1.89)	(0.25)	(1.84)	(0.01)
F norm	8.12	5.20	6.24	6.78	**4.24**	7.20	**49.3**	49.7	49.7	49.4	49.9	49.7
	(0.22)	(0.15)	(0.32)	(0.05)	(0.24)	(0.21)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Specificity	1	1	1	1	1	1	0	0	0	0	0	0
Sensitivity	0.8	**0.99**	0.84	0.93	0.98	0.17	1	1	1	1	1	1
MCC	0.88	**0.93**	0.89	0.92	**0.93**	0.39	0	0	0	0	0	0

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 4.4 *Summary of network signal strength and structure recovery performance across multiple network topologies and estimators for* $p/n = 4$. The median performance values and (standard errors) reported here are based on 30 replications for the BAGENI, BAGENII, BAGL, BAGR, BGHS and GELNET estimators. The best performance value is boldfaced. Cell colours correspond only to the BAGL, BAGR, BAGENI and BAGENII signal strength metrics, with darker colours indicating the best performance value and lighter colours representing the runner-up.

$\frac{p}{n} = 4$	AR(1)						AR(2)					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	67.9	59.8	60.3	57.6	**52.9**	62.7	44.9	31.0	46.7	**29.6**	32.0	30.8
	(0.68)	(0.65)	(1.87)	(0.37)	(1.32)	(0.64)	(0.66)	(0.32)	(2.57)	(0.27)	(1.57)	(0.22)
F norm	15.8	13.3	15.3	13.2	**13.0**	13.4	13.1	7.77	16.3	**7.58**	8.87	7.70
	(0.13)	(0.02)	(0.71)	(0.02)	(0.11)	(0.04)	(0.19)	(0.02)	(1.48)	(0.02)	(0.56)	(0.02)
Specificity	0.44	0.44	0.44	0.43	0.44	0.44	0.96	0.96	**0.97**	0.96	0.96	**0.97**
Sensitivity	0.89	**0.98**	0.84	0.97	0.97	0.83	0.81	**1**	0.6	0.98	0.91	0.57
MCC	0.1	0.1	0.1	0.1	0.1	0.09	**0.46**	**0.46**	0.42	0.44	0.45	0.4
$\frac{p}{n} = 4$	star						circle					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	26.0	**4.62**	39.2	5.45	21.1	10.2	50.1	29.1	37.4	39.5	**28.2**	50.2
	(0.43)	(0.25)	(1.90)	(0.27)	(1.91)	(3.13)	(1.15)	(1.26)	(1.25)	(0.83)	(3.43)	(0.72)
F norm	13.1	**2.28**	22.5	2.47	12.1	2.32	12.3	**4.43**	13.2	5.92	10.2	5.99
	(0.18)	(0.05)	(1.45)	(0.02)	(1.15)	(0.07)	(0.39)	(0.18)	(0.98)	(0.05)	(1.32)	(0.08)
Specificity	0.98	0.98	0.98	0.98	0.98	**0.99**	1	1	1	1	1	1
Sensitivity	0.91	**1**	0.5	**1**	0.87	0.1	0.79	**0.93**	0.74	0.75	0.94	0.12
MCC	0.55	**0.58**	0.4	**0.58**	0.54	0.24	0.84	0.89	0.84	0.79	**0.92**	0.3
$\frac{p}{n} = 4$	scale-free						random					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	23.7	8.33	36.2	9.23	21.3	**6.82**	22.9	15.6	36.5	15.3	22.4	**14.7**
	(0.38)	(0.96)	(2.17)	(0.78)	(2.17)	(0.47)	(0.34)	(1.38)	(1.67)	(1.19)	(1.55)	(1.39)
F norm	12.6	5.25	23.0	5.20	12.8	**5.11**	11.7	6.72	23.5	**6.51**	13.0	6.59
	(0.23)	(0.16)	(1.62)	(0.14)	(1.32)	(0.13)	(0.14)	(0.40)	(1.36)	(0.38)	(1.05)	(0.35)
Specificity	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.97
Sensitivity	0.93	**1**	0.57	**1**	0.93	0.51	0.95	**1**	0.61	**1**	0.92	0.56
MCC	0.55	**0.57**	0.45	**0.57**	0.55	0.43	0.49	0.49	0.42	0.49	0.48	0.42
$\frac{p}{n} = 4$	cluster						hub					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	49.1	32.1	53.3	31.6	37.8	**31.3**	22.8	**15.2**	36.0	15.7	21.2	15.9
	(0.46)	(0.31)	(1.31)	(0.31)	(1.23)	(0.17)	(0.36)	(0.55)	(1.61)	(0.26)	(1.50)	(0.39)
F norm	31.4	29.3	33.5	**29.2**	29.9	29.3	12.9	8.94	23.7	**8.90**	13.5	9.07
	(0.07)	(0.01)	(0.45)	(0.01)	(0.16)	(0.01)	(0.13)	(0.08)	(1.25)	(0.02)	(0.87)	(0.03)
Specificity	0.04	0.04	0.04	0.04	0.04	0.04	0.98	0.98	0.98	0.98	0.98	**0.99**
Sensitivity	1	1	0.99	1	1	0.98	0.94	**1**	0.58	**1**	0.91	0.5
MCC	0.02	0.02	0.02	0.02	0.02	0.02	0.57	**0.58**	0.48	**0.58**	0.57	0.48
$\frac{p}{n} = 4$	band						full					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	**22.9**	25.6	37.0	24.7	23.6	24.7	75.7	64.7	74.3	**61.1**	65.8	64.6
	(0.41)	(0.58)	(1.65)	(0.31)	(1.42)	(0.93)	(0.63)	(0.69)	(1.86)	(0.47)	(1.18)	(0.36)
F norm	11.0	8.96	24.2	**8.82**	13.6	8.96	50.7	**49.6**	50.8	49.7	49.8	49.7
	(0.15)	(0.06)	(1.61)	(0.02)	(1.24)	(0.16)	(0.05)	(0.01)	(0.22)	(0.01)	(0.04)	(0.01)
Specificity	0.98	0.98	**0.99**	0.98	0.98	**0.99**	0	0	0	0	0	0
Sensitivity	0.97	**1**	0.68	**1**	0.93	0.45	1	1	1	1	1	1
MCC	0.6	0.57	0.58	0.57	**0.62**	0.53	0	0	0	0	0	0

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 4.5 *Summary of network signal strength and structure recovery performance across multiple network topologies and estimators for* $p/n = 10$. The median performance values and (standard errors) reported here are based on 30 replications for the BAGENI, BAGENII, BAGL, BAGR, BGHS and GELNET estimators. The best performance value is boldfaced. Cell colours correspond only to the BAGL, BAGR, BAGENI and BAGENII signal strength metrics, with darker colours indicating the best performance value and lighter colours representing the runner-up.

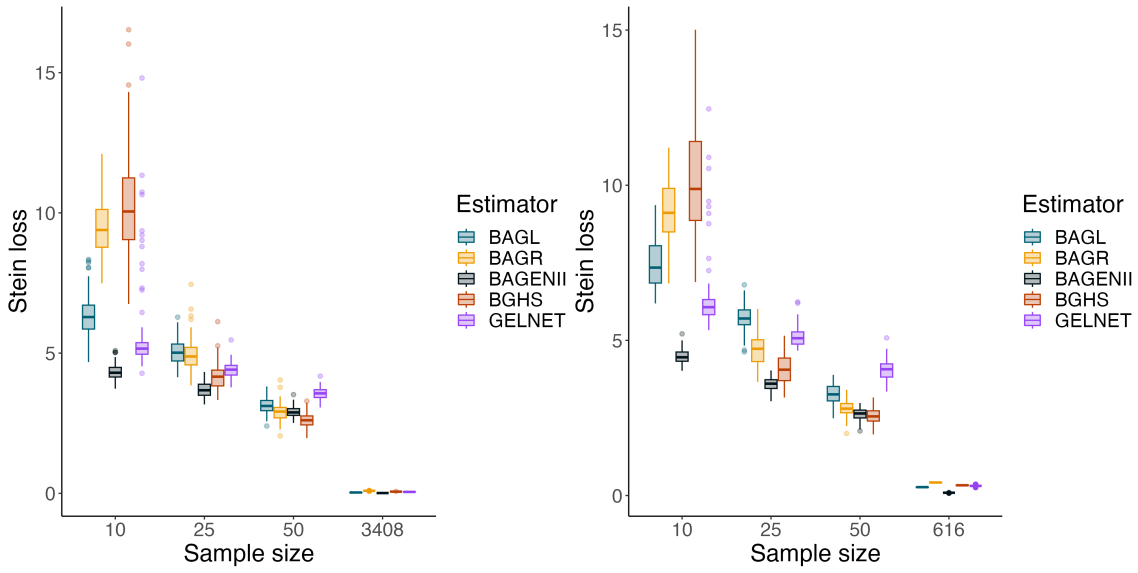
$\frac{p}{n} = 10$	**AR(1)**						**AR(2)**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	83.6	67.3	149	**63.4**	259	64.0	53.3	33.7	119	**31.4**	258	33.1
	(0.37)	(0.42)	(2.25)	(0.23)	(10.9)	(0.72)	(0.37)	(0.27)	(1.96)	(0.19)	(12.8)	(0.66)
F norm	18.2	13.8	68.1	**13.5**	294	13.6	14.7	7.98	65.3	**7.78**	416	7.99
	(0.07)	(0.01)	(2.37)	(0.01)	(44.7)	(0.04)	(0.09)	(0.02)	(2.09)	(0.01)	(69.4)	(0.06)
Specificity	0.43	0.43	0.43	0.43	0.43	**0.44**	0.96	0.96	0.96	0.96	0.96	**0.97**
Sensitivity	0.98	**1**	0.73	**1**	0.76	0.62	0.98	**1**	0.49	**1**	0.44	0.15
MCC	0.09	0.09	0.06	0.09	0.06	0.04	0.44	0.44	0.31	0.44	0.3	0.18
$\frac{p}{n} = 10$	**star**						**circle**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	28.6	5.58	94.0	**4.61**	193	19.6	69.0	**49.4**	108	50.8	113	51.2
	(0.40)	(0.31)	(2.34)	(0.28)	(22.3)	(3.43)	(1.30)	(1.59)	(8.59)	(1.09)	(12.1)	(0.98)
F norm	13.4	2.12	63.7	**1.86**	254	2.64	13.8	6.49	52.2	6.57	61.2	**6.07**
	(0.10)	(0.05)	(2.31)	(0.02)	(59.3)	(0.11)	(0.28)	(0.08)	(5.69)	(0.03)	(11.9)	(0.11)
Specificity	0.98	0.98	0.98	0.98	0.98	**0.99**	0.98	0.98	0.99	0.98	0.99	**1**
Sensitivity	1	1	0.55	1	0.46	0.08	0.87	**0.97**	0.58	0.95	0.74	0.11
MCC	0.57	0.57	0.42	0.57	0.39	0.17	0.64	0.63	0.63	0.58	**0.71**	0.29
$\frac{p}{n} = 10$	**scale-free**						**random**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	25.5	9.23	88.5	**8.31**	257	13.1	24.9	18.1	85.1	**16.5**	252	21.3
	(0.35)	(0.94)	(2.23)	(0.63)	(12.7)	(0.57)	(0.28)	(1.61)	(2.20)	(1.28)	(11.5)	(1.38)
F norm	13.0	5.43	61.5	**5.24**	585	5.63	12.1	**6.78**	62.0	6.87	609	6.98
	(0.25)	(0.19)	(2.30)	(0.16)	(96.8)	(0.10)	(0.12)	(0.38)	(2.14)	(0.36)	(79.7)	(0.30)
Specificity	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.97
Sensitivity	1	1	0.61	1	0.42	0.11	1	1	0.64	1	0.43	0.13
MCC	0.57	0.57	0.45	0.57	0.36	0.17	0.49	0.49	0.41	0.49	0.33	0.18
$\frac{p}{n} = 10$	**cluster**						**hub**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	54.7	33.2	120	**32.0**	277	34.6	24.9	18.9	83.4	**17.1**	247	22.4
	(0.22)	(0.35)	(2.06)	(0.28)	(10.0)	(1.25)	(0.25)	(0.82)	(1.86)	(0.23)	(9.75)	(2.38)
F norm	31.9	**29.2**	69.9	29.3	524	29.3	13.4	9.30	60.2	**9.13**	577	9.33
	(0.03)	(0.01)	(1.90)	(0.01)	(69.9)	(0.03)	(0.09)	(0.08)	(2.09)	(0.01)	(73.8)	(0.04)
Specificity	0.04	0.04	0.04	0.04	0.04	0.04	0.98	0.98	0.98	0.98	0.98	0.98
Sensitivity	1	1	0.99	1	0.98	0.96	1	1	0.61	1	0.41	0.12
MCC	0.02	0.02	0.02	0.02	0.02	0	0.58	0.58	0.46	0.58	0.38	0.21
$\frac{p}{n} = 10$	**band**						**full**					
	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET	BAGENI	BAGENII	BAGL	BAGR	BGHS	GELNET
Stein	**25.3**	31.3	81.1	28.2	239	33.0	85.1	66.4	151	**63.7**	285	66.4
	(0.27)	(0.55)	(2.10)	(0.23)	(9.77)	(0.87)	(0.38)	(0.50)	(1.80)	(0.51)	(11.0)	(1.61)
F norm	11.4	9.62	61.2	**9.33**	578	9.41	51.2	49.8	82.1	**49.7**	387	49.8
	(0.10)	(0.04)	(2.30)	(0.01)	(77.3)	(0.05)	(0.04)	(0.01)	(1.52)	(0.01)	(55.2)	(0.02)
Specificity	0.98	0.98	0.98	0.98	0.98	**0.99**	0	0	0	0	0	0
Sensitivity	1	1	0.68	1	0.47	0.14	1	1	1	1	1	1
MCC	0.57	0.57	0.49	0.57	0.42	0.25	0	0	0	0	0	0

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

4.5 SEER breast cancer

The performance of the estimators considered in the synthetic studies in Section 4.4 is evaluated on breast cancer data from the Surveillance, Epidemiology, and End Results (SEER) program, managed by the National Cancer Institute (NCI), and is based on the November 2017 update. The data set is publicly available in (Teng, 2019). It specifically pertains to female patients diagnosed with infiltrating duct and lobular carcinoma breast cancer (identified by SEER primary cites recode NOS histology codes 8522/3) within the 2006-2010 time frame. Data selection procedures ensured exclusion of patients with indeterminable tumour size, unknown status of regional Lymph Nodes (LNs), unconfirmed number of positive regional LNs, and those with survival duration less than a month. As a result, the final dataset encompasses 15 attributes from 4024 patients. The attributes include continuous, categorical and binary measurements, the latter two are categorically encoded resulting in 32 continuous measurements per patient. Following the approach of (Tang et al., 2020), the data is standardised using a non-paranormal transformation in order to satisfy the Gaussian assumption. The attributes contain socio-demographic information, pathological cancer characteristics and the survival status of the patient. Descriptions of each attribute are available in Subsection C.1.1 of the appendix. The dataset is split into two distinct groups according to the patients' mortality status - those who survived ($n = 3408$) and those who succumbed to the disease ($n = 616$).

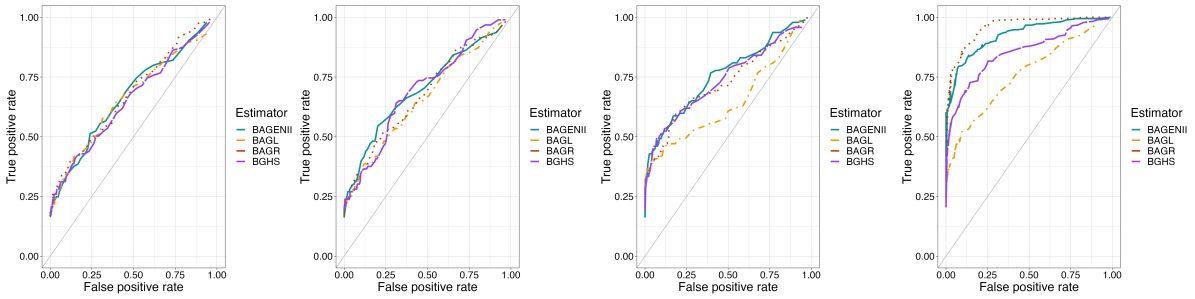


(a) alive (b) deceased

Fig. 4.2 *Loss comparison between the BAGL, BAGR, BAGENII, BGHS and GELNET on the SEER breast cancer data.* The left-hand panel illustrates Stein's loss on the cohort of patients who survived the disease and the right-hand panel depicts Stein's loss for the the cohort of patients who succumbed to the disease.

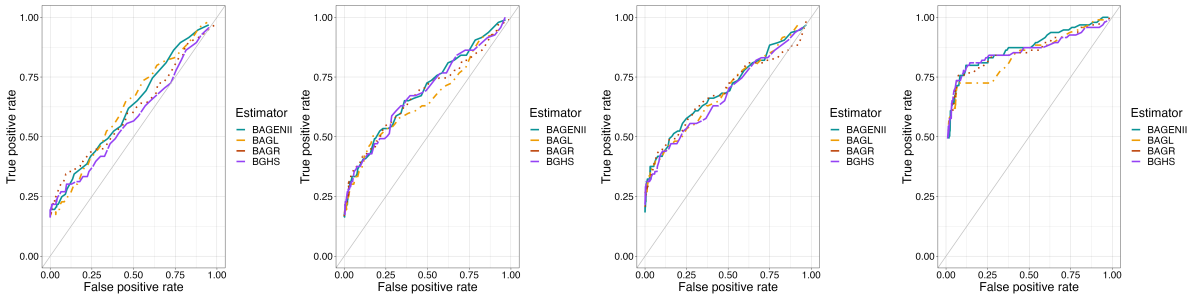
The cancer data is clearly not high-dimensional where $p > n$. Similarly to the approach in (Wieringen and Peeters, 2016), high-dimensionality can also be assessed here by subsampling. In this scenario the sample sizes considered for both cohorts are $n = 10$, 25 and 50. The full sample size is also considered. The precision estimates are evaluated in terms of Stein's loss and the Frobenius norm (as defined in Section 4.4). Clearly these assessments require the population precision matrix which is unknown. As a proxy, Stein's estimator (Haff, 1991) of $\boldsymbol{\Omega}$ using the full sample space for each cohort is used. The results of the loss evaluation are displayed in Figure 4.2 and

© University of Pretoria



(a) $n = 10$ (b) $n = 25$ (c) $n = 50$ (d) $n = 3408$

Fig. 4.3 *Receiver operating characteristic curves for the BAGL, BAGR, BAGENII, BGHS and GELNET on the SEER breast cancer data for surviving patients.* The curves investigate the performance of the estimator's network recovery abilities for various p/n ratios.



(a) $n = 10$ (b) $n = 25$ (c) $n = 50$ (d) $n = 616$

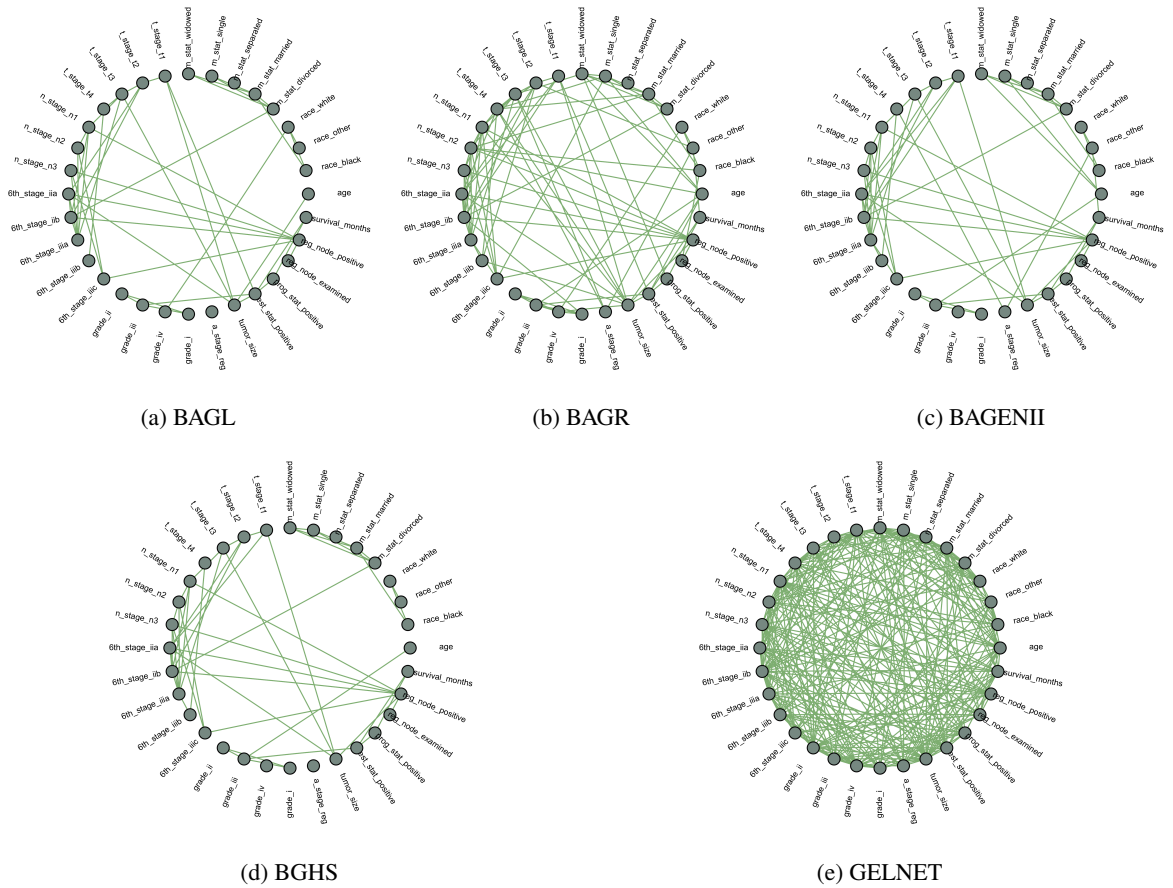
Fig. 4.4 *Receiver operating characteristic curves for the BAGL, BAGR, BAGENII, BGHS and GELNET on the SEER breast cancer data for deceased patients.* The curves investigate the performance of the estimator's network recovery abilities for various p/n ratios.

C.1 in Section C of the appendix. These figures show that the BAGENII estimator enjoys the adaptability of its constituent building blocks, in that the loss (Stein and Frobenius) associated with this estimator is amongst the lowest across all estimators considered. As to be expected, all estimators yield lower loss values for larger sample sizes.

Regarding structure learning, Figures 4.3 and 4.4 display the receiver operating characteristic (ROC) curves for each estimator, sample size and cohort considered. The ROC curves for the Bayesian estimators are generated by varying the width of symmetric credible interval from 1% to 99%. Again, it is clear that all estimators perform better when n is larger than p, however, this performance deteriorates considerably under the compliment setting. The networks in Figures 4.5 and C.2 are constructed using the full data set and 90% credible interval on each $\omega_{i\neq j}$. The true topology is not known, however, the proxy adjacency matrix is determined if $|\omega_{i\neq j}| > 0.05$ under Stein's estimator. Focussing on the deceased cohort in Figure 4.5, the BAGL, BAGENII and BGHS graphs display similar topological representations across both cohorts. The BAGR estimator provides the densest graph of all the Bayesian estimators, however, the GELNET is practically fully connected and interpreting the connectivity here may prove challenging. The networks for the deceased patients are visibly sparser compared to the that of the surviving patients - see Figure C.2. These networks could provide insights into the interconnections among biomarkers and demographic attributes in the complex landscape of breast cancer.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



(a) BAGL (b) BAGR (c) BAGENII

(d) BGHS (e) GELNET

Fig. 4.5 *The inferred networks from the BAGL, BAGR, BAGENII, BGHS and GELNET for the SEER breast cancer data for deceased patients.* The Bayesian networks are determined using a 90% symmetric credible interval.

4.6 Conclusion

The naïve Bayesian graphical elastic net prior and its adaptive variant are introduced for the estimation of precision matrices under the assumption of Gaussianity. The form of the prior for $\mathbf{\Omega}$ makes use of the product of double exponential and Gaussian for the off-diagonal elements and the product of exponential and truncated Gaussian for main diagonal elements. The estimator is implemented via a block Gibbs sampler. The posterior mode resulting from this prior contains both an ℓ_1 and ℓ_2 constraint. Synthetic simulations illustrate the ability of the naïve Bayesian adaptive graphical elastic net to offer a balanced performance in the presence of varying sparsity and p/n ratios. Structure learning is achieved through a heuristic treatment of the symmetric credible interval based on the posterior MCMC chains of $\mathbf{\Omega}$. The application of the proposed estimator to data encompassing breast cancer, suggest that the estimator has the adaptability to navigate complex multidimensional applications under varying unknown sparsity conditions. The naïve Bayesian adaptive graphical elastic net is also implemented in the 'baygel' R-package which is freely available from the Comprehensive R Archive Network.

It is clear that probabilistic networks are becoming ever-present in a multitude of scientific disciplines. That being said, there is growing interest in the scenario where networks may undergo changes over time or where they may differ between states. Identifying and studying the changes between two networks from a Bayesian perspective is the focus of the next and final chapter.

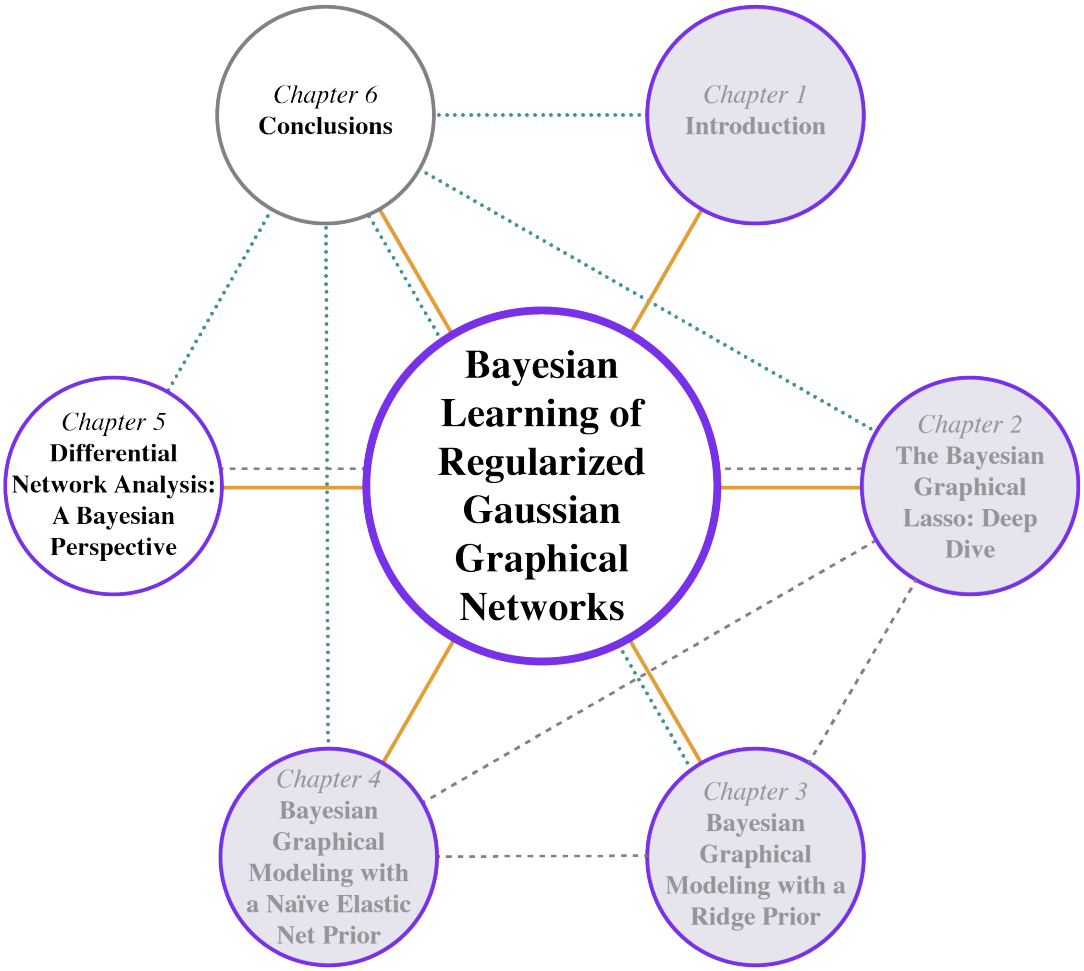
© University of Pretoria

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chapter 5

Differential Network Analysis: A Bayesian Perspective



© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

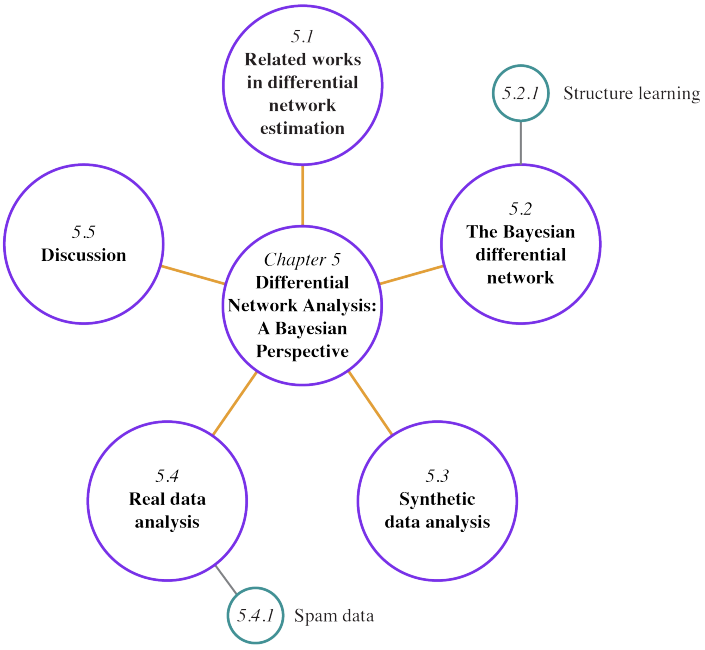


Fig. 5.1 *Chapter 5 visual outline.*

Differential networks (DN) are important tools for capturing the changes in conditional (in)dependencies between two samples. A Bayesian approach for estimating DNs is introduced. The Bayesian approach separately estimates the components of the DN using the set of adaptive priors introduced in Chapter 2. Synthetic experiments illustrate strong performance of Bayesian DN in numerical accuracy and graphical structure determination when compared to state of the art methods. The proposed Bayesian method is applied to South African COVID-19 data to investigate the change in DN structure between various phases of the pandemic. It is worth mentioning that the work presented here was completed as the initial component of this dissertation. That being said, the Bayesian methodologies introduced in Chapters 3 and 4 are not utilised here, however, they offer promising avenues for future exploration to enhance the methods presented in this chapter. Figure 5.1 provides a visual outline of this chapter.

5.1 Related works in differential network estimation

The work presented in this chapter hinges on a fundamental assumption: that the observations, $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{n_1}$ and $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{n_2}$, are generated from multivariate Gaussian distributions, $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Omega}_1^{-1})$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Omega}_2^{-1})$, respectively. Here, n_1 and n_2 represent the respective sample sizes, which need not be equal. The central focus of this analysis lies in the estimation of the DN, denoted as $\boldsymbol{\Delta} = \boldsymbol{\Omega}_2 - \boldsymbol{\Omega}_1$, encapsulating the difference between two precision matrices. Recall from Chapter 1 that statistical techniques for estimating $\boldsymbol{\Delta}$ are on the rise. These techniques can largely be classified into two main categories. The first estimating the individual precision matrices, $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ separately; where the estimated DN is the difference between the estimated precision matrices. For example, the methods and references for GGM estimation outlined in Chapters 1 - 4 can be used to directly estimate $\boldsymbol{\Delta}$. The second methodology estimates both the precision matrices simultaneously. The approach here, typically penalises

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

a joint loss function for both precision matrices. (Chiquet et al., 2011) provide a methodology for inference and estimation of functions of GGMs. In particular, the Intertwined Graphical Lasso (IGL) approach biases the estimation of the precision matrices towards a common value. More so, their Graphical Cooperative Lasso (GCL) utilises a group-penalty for solutions that favour a common sparsity pattern. (Guo et al., 2011) and (Zhu and Li, 2018) estimate separate graphical models using a joint penalised loss function. (Zhao et al., 2014) propose a method for estimating $\mathbf{\Delta}$ directly which relaxes the need for both individual precision matrices to be sparse nor be estimated directly. Similarly, (Yuan et al., 2017) and (Jiang et al., 2018) utilise an alternating direction method of multipliers (ADMM) algorithm for estimating $\mathbf{\Delta}$ from their joint ℓ_1 penalised convex loss function. More recently, (Tang et al., 2020) introduce a computationally efficient iterative shrinkage-thresholding algorithm for minimising the ℓ_1 loss function defined in (Jiang et al., 2018), namely

$$L_1(\mathbf{\Delta}) = \frac{1}{2}\text{trace}(\mathbf{\Delta}^\top \mathbf{S_1}\mathbf{\Delta}\mathbf{S_2}) - \text{trace}(\mathbf{\Delta}(\mathbf{S_1} - \mathbf{S_2})), \tag{5.1}$$

is convex and $\mathbf{S}_1$ and $\mathbf{S}_2$ are the sample covariance matrices. The DN estimate is obtained by minimising the penalised loss in (5.1). An analogous symmetric convex loss function and estimator is proposed by (Yuan et al., 2017)

$$\hat{\mathbf{\Delta}}_1 = \underset{\mathbf{\Delta}\in\mathbb{R}^{p\times p}}{\arg\min}\, \frac{1}{2}\text{trace}(\mathbf{\Delta}^\top \mathbf{S_1}\mathbf{\Delta}\mathbf{S_2}) - \text{trace}(\mathbf{\Delta}(\mathbf{S_1} - \mathbf{S_2})) + \lambda\|\mathbf{\Delta}\|_1, \tag{5.2}$$

where $\|\mathbf{\Delta}\|_1 = \sum_{1\leq i,\, j\leq p} |\omega_{2:ij} - \omega_{1:ij}|$ represents the ij'th entry of the k'th precision matrix. The estimator in (5.2) is not symmetric and (Yuan et al., 2017) proposed the following symmetric convex loss function

$$L_2(\mathbf{\Delta}) = \frac{L_1(\mathbf{\Delta}) + L_1(\mathbf{\Delta}^\top)}{2}, \tag{5.3}$$

and a symmetric estimator

$$\hat{\mathbf{\Delta}}_2 = \underset{\mathbf{\Delta}\in\mathbb{R}^{p\times p}}{\arg\min}\, \frac{1}{4}\left(\text{trace}(\mathbf{\Delta}^\top \mathbf{S_1}\mathbf{\Delta}\mathbf{S_2}) + \text{trace}(\mathbf{\Delta}^\top \mathbf{S_2}\mathbf{\Delta}\mathbf{S_1}\right) - \text{trace}(\mathbf{\Delta}(\mathbf{S_1} - \mathbf{S_2})) + \lambda\|\mathbf{\Delta}\|_1. \tag{5.4}$$

Finally, (Yuan et al., 2017) and (Jiang et al., 2018) illustrate that $\hat{\mathbf{\Delta}}_1$ and $\hat{\mathbf{\Delta}}_2$ are consistent estimators of $\mathbf{\Delta}$.

The shrinkage-thresholding algorithm proposed by (Tang et al., 2020), based on the fast-iterative shrinkage-thresholding algorithm in (Beck and Teboulle, 2009), aims to minimise (5.1). The objective function is given by

$$\underset{\mathbf{\Delta}\in\mathbb{R}^{p\times p}}{\arg\min}\, L_1(\mathbf{\Delta}) + \rho\|\mathbf{\Delta}\|_1. \tag{5.5}$$

The lasso tuning parameter, ρ, controls the strength of the penalty term and resultantly the amount of shrinkage applied to the precision matrix entries. The optimisation objective converges to the solution sequentially using a quadratic approximation and a gradient descent algorithm. The efficiency of the procedure is attested to this approach, resulting in superior computational complexity in contrast to the ADMM approaches by (Yuan et al., 2017) and (Jiang et al., 2018). To conclude this section, it is worth noting that the iterative shrinkage-thresholding method is used as the frequentist approach in the synthetic and application studies presented in Sections 5.3 and 5.4.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

5.2 The Bayesian differential network

A Bayesian treatment of DNs remains, to our knowledge, unexplored, and the methodology introduced here aims to develop a Bayesian estimation procedure in which each component of the DN is separately estimated using Bayesian precision estimation techniques rooted in absolutely continuous priors. The framework utilises the Bayesian adaptive graphical lasso from Section 2.4 in Chapter 2 as a starting point. To this end, the remainder of this section briefly revisits the Bayesian graphical lasso for the sake of readability. As mentioned in the opening statement, the Bayesian adaptive graphical ridge-type and naïve elastic net priors from Section 3.4 in Chapter 3 and Section 4.3 in Chapter 4 can also be employed in a 'plug and play' manner, so to speak.

Recall from Section 2.1 in Chapter 2 that the Bayesian graphical lasso is founded upon a prior from which the posterior mode results in the objective function of the graphical lasso. The objective of the graphical lasso is to maximise the penalised log-likelihood

$$\underset{\boldsymbol{\Omega}\in\mathbb{M}^{+}}{\arg\max}\left\{\log(\det\boldsymbol{\Omega})-\operatorname{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right)-\rho_{GL}\|\boldsymbol{\Omega}\|_{1}\right\}. \tag{5.6}$$

Here, $\rho_{BGL}\geq 0$ is the shrinkage parameter, where the subscript 'GL' is purely aesthetic and serves to distinguish the notation used in the iterative shrinkage-thresholding objective function in (5.5). The Bayesian connection to the graphical lasso is the MAP estimation of the following prior

$$\begin{aligned} p\left(\mathbf{y}_{i}\,|\,\boldsymbol{\Omega}\right)&=\mathrm{N}\left(\mathbf{y}_{i}\,|\,\mathbf{0},\boldsymbol{\Omega}^{-1}\right)\quad(i=1,\ldots,n)\\ p\left(\boldsymbol{\Omega}\,|\,\lambda\right)&=C^{-1}\prod_{i<j}\left\{\mathrm{DE}(\omega_{ij}\,|\,\lambda)\right\}\prod_{i=1}^{p}\left\{\mathrm{EXP}(\omega_{ii}\,|\,\frac{\lambda}{2})\right\}\mathbb{1}_{\boldsymbol{\Omega}\in\mathbb{M}^{+}}. \end{aligned} \tag{5.7}$$

The corresponding posterior is given by

$$p\left(\boldsymbol{\Omega}\,|\,\mathbf{Y},\lambda\right)\propto\det\boldsymbol{\Omega}^{\frac{n}{2}}\exp\left\{-\frac{1}{2}\operatorname{trace}(\mathbf{S}\boldsymbol{\Omega})-\sum_{i\neq j}\frac{\lambda}{2}|\omega_{ij}|-\sum_{i=1}^{p}\frac{\lambda}{2}\omega_{ii}\right\}.$$

The log of the posterior distribution is given by

$$p_{\ell}\left(\boldsymbol{\Omega}\,|\,\mathbf{Y},\lambda\right)\propto\log(\det\boldsymbol{\Omega})-\operatorname{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right)-\sum_{i=1}^{p}\frac{\lambda}{n}\omega_{ii}-\sum_{i\neq j}\frac{\lambda}{n}|\omega_{ij}|. \tag{5.8}$$

The MAP estimate is the value of $\boldsymbol{\Omega}$ which maximises (5.8), namely

$$\hat{\boldsymbol{\Omega}}=\underset{\boldsymbol{\Omega}\in\mathbb{M}^{+}}{\arg\max}\left\{\log(\det\boldsymbol{\Omega})-\operatorname{trace}\left(\frac{\mathbf{S}}{n}\boldsymbol{\Omega}\right)-\rho_{BGL}\|\boldsymbol{\Omega}\|_{1}\right\},$$

where $\rho_{BGL}=\lambda/n$ and within the Bayesian context λ is treated as the shrinkage parameter. The subscript 'BGL' is again purely aesthetic and serves to distinguish the notation used in the iterative shrinkage-thresholding and grahpical lasso objective function in (5.5) and (5.6), respectively.

The estimation of $\boldsymbol{\Omega}$ is achieved through the use of a block Gibbs sampler, which requires an extension of the prior in (5.7). Conveniently, the aforementioned prior shares archetypal characteristics with the Bayesian lasso in the regression domain presented in (Park and Casella, 2008). The construction of the Gibbs sampler for the Bayesian lasso is derived from a heirarchical prior in which the double exponential distribution is represented as a scale mixture of Gaussians, (Andrews and Mallows, 1974; West, 1987). This formulation is grounded in the

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

assumption of independence of the conditional double exponential priors which is not up held in the Bayesian graphical lasso setting due to the positive definite constraint on $\boldsymbol{\Omega}$. The prior in (5.7) can be reformulated as

$$\begin{aligned} p\left(\mathbf{y}_i \,|\, \boldsymbol{\Omega}\right) &= \mathrm{N}\left(\mathbf{y}_i \,|\, \mathbf{0}, \boldsymbol{\Omega}^{-1}\right) \quad (i = 1, \ldots, n) \\ p\left(\boldsymbol{\Omega} \,|\, \boldsymbol{\tau}, \lambda\right) &= C_{\tau}^{-1} \prod_{i<j} \left\{ \frac{1}{\sqrt{2\pi\tau_{ij}}} \exp\left(-\frac{\omega_{ij}^2}{2\tau_{ij}} \right) \right\} \prod_{i=1}^{p} \left\{ \frac{\lambda}{2} \exp\left(-\frac{\lambda}{2}\omega_{ii} \right) \right\} \mathbb{1}_{\boldsymbol{\Omega} \in \mathbb{M}^+} \\ p(\boldsymbol{\tau} \,|\lambda) &\propto C_{\tau} \prod_{i<j} \frac{\lambda^2}{2} \exp\left(-\frac{\lambda^2}{2}\tau_{ij} \right), \end{aligned} \tag{5.9}$$

where $\boldsymbol{\tau} = \{\tau_{ij}\}_{i<j}$ is a vector of the scale parameters. The normalising constant, C_{τ}, has no closed-form solution but the techniques described in Subsection 2.1.1 in Chapter 2 can be used to circumvent the intractable constants. Finally, the posterior distribution from which the Bayesian graphical lasso Gibbs sampler generates $\boldsymbol{\Omega}$ is given by

$$\begin{aligned} p(\boldsymbol{\Omega}, \tau \,|\, \mathbf{Y}, \lambda) \propto \det\boldsymbol{\Omega}^{\frac{n}{2}} \exp\left\{ -\operatorname{trace}\left(\frac{1}{2}\mathbf{S}\boldsymbol{\Omega} \right) \right\} \prod_{i<j} \left\{ \tau_{ij}^{-\frac{1}{2}} \exp\left(-\frac{\omega_{ij}^2}{2\tau_{ij}} \right) \exp\left(-\frac{\lambda^2}{2}\tau_{ij} \right) \right\} \\ \times \prod_{i=1}^{p} \left\{ \exp\left(-\frac{\lambda}{2}\omega_{ii} \right) \right\} \mathbb{1}_{\boldsymbol{\Omega} \in \mathbb{M}^+}. \end{aligned} \tag{5.10}$$

The form specified above requires the selection of an appropriate λ. The hierarchical form in (5.9) and the block Gibbs sampler in Algorithm 1 may be tailored to allow for adaptive shrinkage on the off-diagonal elements ω_{ij}. The Bayesian adaptive graphical lasso assigns a set of gamma priors to $\{\lambda_{ij}\}_{i \leq j}$

$$\begin{aligned} p(\boldsymbol{\Omega} \,|\, \{\lambda_{ij}\}_{i \leq j}) &= C_{\{\lambda_{ij}\}_{i \leq j}}^{-1} \prod_{i<j} \left\{ \mathrm{DE}(\omega_{ij} | \lambda_{ij}) \right\} \prod_{i=1}^{p} \left\{ \mathrm{EXP}(\omega_{ii} | \frac{\lambda_{ii}}{2}) \right\} \mathbb{1}_{\boldsymbol{\Omega} \in \mathbb{M}^+} \\ p(\{\lambda_{ij}\}_{i<j} \,|\, \{\lambda_{ii}\}_{i=1}^{p}) &\propto C_{\{\lambda_{ij}\}_{i \leq j}} \prod_{i<j} \mathrm{GA}(r, s). \end{aligned} \tag{5.11}$$

These technical preliminaries describe the building blocks upon which the Bayesian DN is devised.

5.2.1 Structure learning

Recall that, under the assumption of Gaussianity, the precision matrix directly determines the conditional (in)dependence relations of the attributes under investigation. The magnitude of the values in the precision matrix describes the strength of this association, while the binarised partial correlation matrix (a function of the precision matrix) can be used to encode the presence of an association, typically represented graphically as a network composed of vertices and edges. Structure learning or network recovery for the Bayesian DN requires a heuristic treatment since the Bayesian graphical lasso is a shrinkage method rooted in absolute continuous priors and therefore does not estimate off-diagonal elements of $\boldsymbol{\Omega}$ to be exactly equal to zero - see Chapters 2 - 4 for more details. A momentary departure from the prevailing use of present tense in the writing style is necessary to provide clarity. In the remainder of this Subsection, the emphasis is placed on adaptations of the Wishart thresholding approach described in Chapter 2. The thresholding approach served as the initial point of entry into heuristic structure learning in the entirety of this dissertation, and the exploration of the symmetric credible interval approach, although discussed in earlier chapters, followed later on. With that said, considering the heuristic nature of these structure learning techniques, the procedure described in this chapter remains agnostic to the chosen structure learning technique.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The network recovery for the Bayesian graphical lasso presented in the seminal work of (Wang, 2012) asserts that $\omega_{ij} \neq 0$ if

$$\frac{\tilde{p}_{ij}}{E_g(p_{ij} \mid \mathbf{Y})} > \eta. \tag{5.12}$$

Here, $\eta \in [0,1]$ and originally set to 0.5. Next, $\tilde{p}_{ij}$ is the posterior sample mean estimate of the partial correlation associated with the Bayesian graphical lasso prior and g is the standard conjugate Wishart prior $W(3, \mathbf{I}_p)$. The degrees of freedom parameter is selected based on the simulated experiments and gene expression application in (Jones et al., 2005). It is not unexpected that as the attribute space dimensionality grows, so does the diversity of network topologies, each exhibiting varying levels of sparsity. It is therefore of interest to examine whether a recommendation for η based on topology is more appropriate. Resultantly, a simple thresholding criterion which is agnostic to the underlying precision matrix estimation technique is proposed, and referred to as Smith's for labelling purposes in subsequent analyses albeit a simple Wishart decision theoretic approach.

$$|E_h(p_{ij} \mid \mathbf{Y})| > \eta. \tag{5.13}$$

Here, h is the standard conjugate Wishart $\mathrm{W}(3, \epsilon \mathbf{I}_p)$ described in (Williams, 2021) and ϵ a constant. Details on the computation of the posterior modes in (5.12) and (5.13) can be found in Section 2.2 of Chapter 2.

The forthcoming synthetic data analysis section describes the simulation procedure (akin to those presented in previous chapters) designed to assess the performance of the Bayesian DN under varying network topologies and association strengths. The goal here, however, is to suggest a suitable sparsity threshold region under the varying network topologies for the recommended sparsity criterion in (5.13). The Bayesian DN is employed across all network topologies, using thresholds denoted as η that fall within the range of 0.01 to 0.99, with increments of 0.02. This procedure is repeated 40 times, for $p \in \{10, 30, 100\}$, so that all metrics displayed are based on measures of center statistics, in this case the mean.

The receiver operating characteristic curves (ROC) under varying η for each Bayesian sparsity criterion in (5.12) and (5.13) are displayed in Figure 5.3. It is clear that the thresholding techniques exhibit similar characteristics, however, the Smith's technique in (5.13) is arguably marginally better in terms of the trade off between the true positive and false positive rate (TPR and FPR), respectively. Next, Figure 5.2 illustrates how the F1 score, per network topology, is impacted under varying thresholds η. There is no clear golden threshold across the topologies. This is unsurprising since the topologies exhibit different sparsity levels. That being said, three observations are noteworthy. First, the boundary thresholds rarely produce the highest F1 score for Smith's threshold technique. In fact, the best performing threshold generally appears to occur in the region $\eta^* \in \{0.2 - 0.4\}$. Second, across all dimensions considered, Smith's threshold technique obtains the highest F1 scores. Lastly, the standard error across both threshold techniques and threshold regions is relatively stable. Both threshold techniques perform comparably well noting, however, that Smith's method requires less computation. The figures for $p \in \{30, 100\}$ are available in Section D.1 of the appendix.

5.3 Synthetic data analysis

The synthetic experiment is designed to test the parameter estimation and graphical structure determination of the DN estimation for both the novel Bayesian approach (referred to as 'B-net') and the iterative shrinkage-thresholding estimator (referred to as 'D-net') from (Tang et al., 2020). The iterative shrinkage-thresholding estimator uses the lasso penalty and Bayesian Information Criterion (BIC) for model estimation and selection, respectively. For all

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

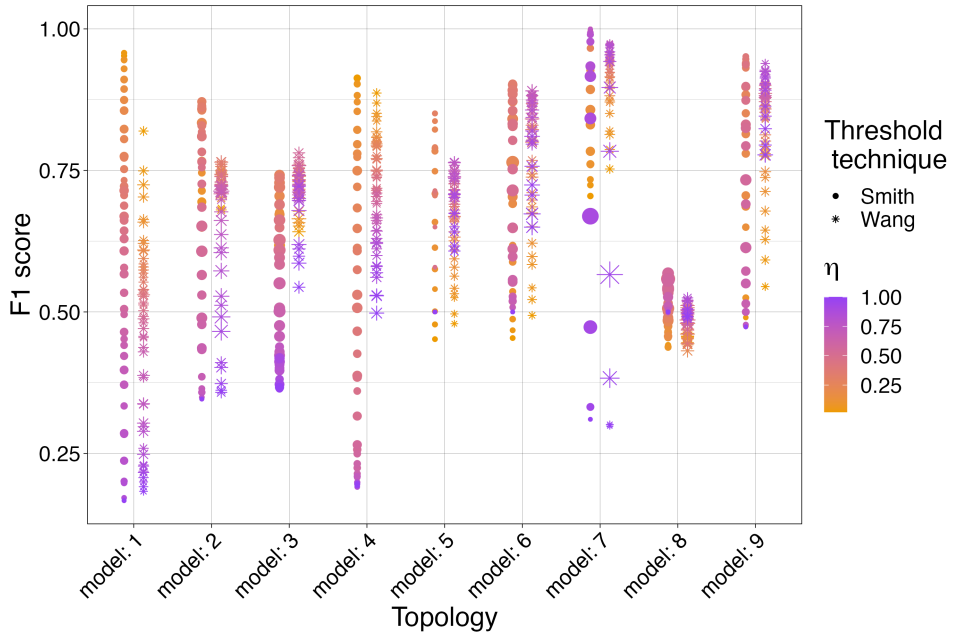


Fig. 5.2 *F1 score for various topologies, thresholding techniques and varying thresholds.* The curves display the network recovery performance according to the F1 score for various graph structures under varying thresholds for each Bayesian sparsity criterion in (5.12) and (5.13) for dimension $p = 10$. The size of the point is proportional to the standard error obtained during simulation.

simulations, the assumption is that the observations, $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{n_1}$ and $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{n_2}$ are generated from a Gaussian $N_p(\mathbf{0}, \mathbf{\Omega}_1^{-1})$ and $N_p(\mathbf{0}, \mathbf{\Omega}_2^{-1})$ respectively. The true DN is

$$\mathbf{\Delta} = \mathbf{\Omega}_2 - \mathbf{\Omega}_1,$$

where the true precision matrices are $\mathbf{\Omega}_1 = \mathbf{\Sigma}_1^{-1}$ and $\mathbf{\Omega}_2 = \mathbf{\Sigma}_2^{-1}$. The sample sizes and dimensions for each model are $n_1 = n_2 \in \{50, 100, 200\}$ and $p_1 = p_2 \in \{10, 30, 100\}$, respectively. The Bayesian estimates are based on 10000 Monte Carlo iterations after 5000 burn-in iterations.

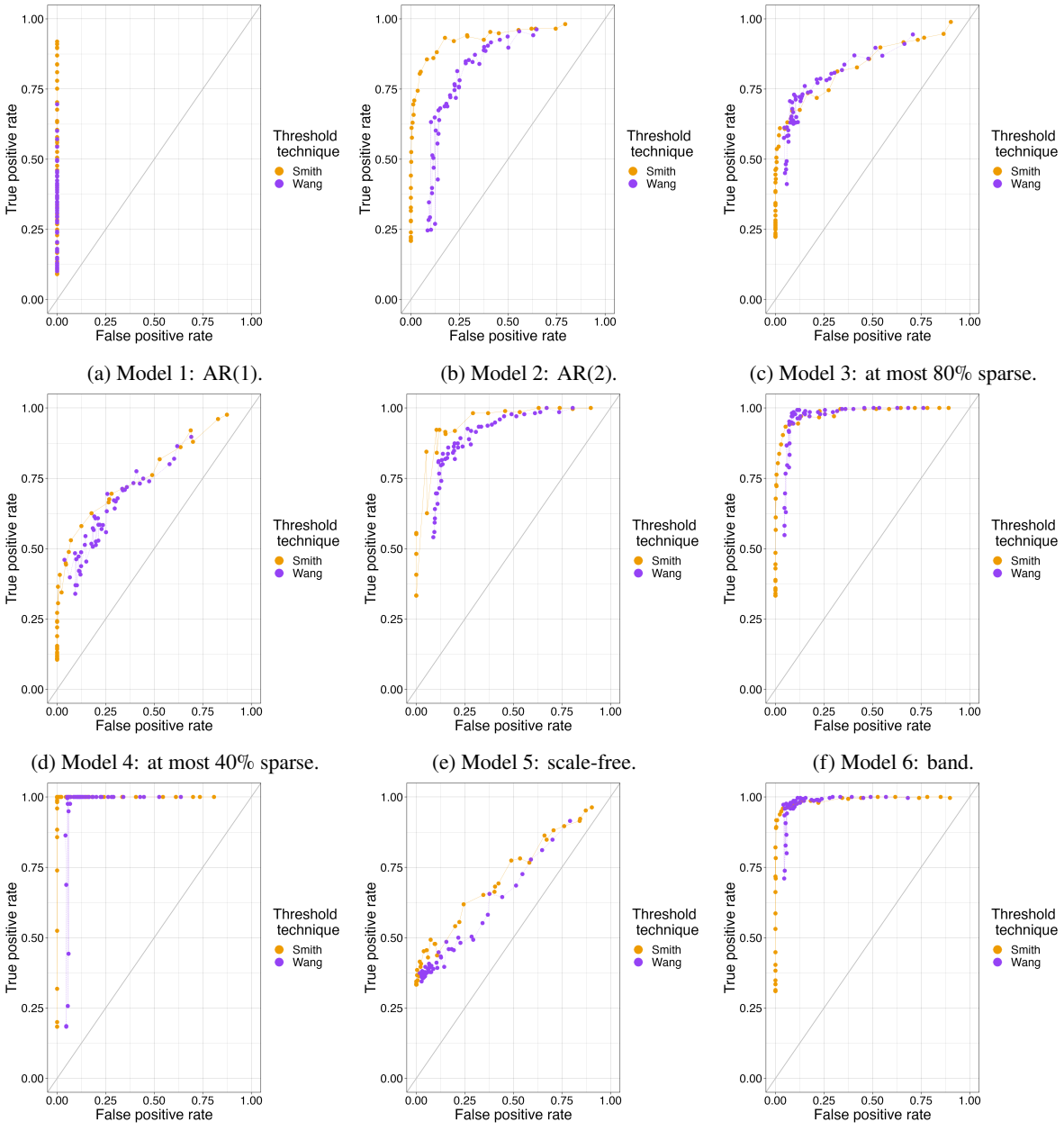
Model configuration. The Bayesian DN applies the Bayesian adaptive graphical lasso (BAGL) in (5.11) in the estimation of the components in $\mathbf{\Delta}$. Here, the parameters of the gamma priors in the BAGL estimator are set to $r = 10^{-2}$ and $s = 10^{-6}$, as proposed by the founding author, for the hyperparameters of the prior distributions of λ_{ij} for $i < j$ and $\lambda_{ii} = 1$ for $i = 1, \ldots, p$. The iterative shrinkage-thresholding approach jointly estimates the precision matrices by minimising (5.1).

Simulation configuration. The following nine graphical structure variations (topologies) are considered - where the structure of each is applied to each component in the DN's composition to achieve the desired structure in the DN itself - in the simulation. It is worth noting that the topologies considered here are similar, but need not be identical to those considered in Sections 3.5 and 4.4 within Chapters 3 and 4, respectively. What is clear here, is the simplying assumption that the components of each synthetic DN considered have identical topological structures.

1. *topology 1*: an AR(1) model.
   - $\mathbf{\Omega}_1$: $\omega_{ij} = 0.7^{|i-j|}$.
   - $\mathbf{\Omega}_2$: $\omega_{ij} = 0.75^{|i-j|}$.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



(a) Model 1: AR(1). (b) Model 2: AR(2). (c) Model 3: at most 80% sparse.

(d) Model 4: at most 40% sparse. (e) Model 5: scale-free. (f) Model 6: band.

(g) Model 7: cluster. (h) Model 8: star. (i) Model 9: circle.

Fig. 5.3 *Receiver operating characteristic curves for various topologies thresholding techniques.* The ROC curves display the network recovery performance for various graph structures under varying thresholds for each Bayesian sparsity criterion in (5.12) and (5.13) for dimension $p = 10$.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

2. *topology 2*: an AR(2) model.
   - $\mathbf{\Omega}_1$: $\omega_{ii} = 0.1, \omega_{i,i-1} = \omega_{i-1,i} = 0.05$ and $\omega_{i,i-2} = \omega_{i-2,i} = 0.025$.
   - $\mathbf{\Omega}_2$: $\omega_{ii} = 1, \omega_{i,i-1} = \omega_{i-1,i} = 0.5$ and $\omega_{i,i-2} = \omega_{i-2,i} = 0.25$.
3. *topology 3*: a sparse random model where both components have approximately up to 80% off-diagonal elements set to zero.
4. *topology 4*: a moderately sparse random model where both components have approximately up to 40% off-diagonal elements set to zero.
5. *topology 5*: a scale-free model where the second component is a scalar multiple of the first.
6. *topology 6*: a band or diagonal model.
   - $\mathbf{\Omega}_1$: $\omega_{ii} = 1, \omega_{ij} = 0.2$ for $1 \leq i \neq j \leq p/2$, $\omega_{ij} = 0.5$ for $p/2+1 \leq i \neq j \leq p$ and $\omega_{ij} = 0$ otherwise.
   - $\mathbf{\Omega}_2$: $\omega_{ii} = 1, \omega_{ij} = 0.7$ for $1 \leq i \neq j \leq p/2$, $\omega_{ij} = 0.9$ for $p/2+1 \leq i \neq j \leq p$ and $\omega_{ij} = 0$ otherwise.
7. *topology 7*: a cluster model containing two disjoint groups.
   - $\mathbf{\Omega}_1$: $\omega_{ii} = 1, \omega_{ij} = 0.5$ for $1 \leq i \neq j \leq p/2$, $\omega_{ij} = 0.5$ for $p/2+1 \leq i \neq j \leq p$ and $\omega_{ij} = 0$ otherwise.
   - $\mathbf{\Omega}_2$: $\omega_{ii} = 1, \omega_{ij} = 0.9$ for $1 \leq i \neq j \leq p/2$, $\omega_{ij} = 0.9$ for $p/2+1 \leq i \neq j \leq p$ and $\omega_{ij} = 0$ otherwise.
8. *topology 8*: a star model with every node connected to the first node.
   - $\mathbf{\Omega}_1$: $\omega_{ii} = 1, \omega_{1,i} = \omega_{i,1} = 0.1$ and $\omega_{i,j} = 0$. otherwise.
   - $\mathbf{\Omega}_2$: $\omega_{ii} = 1, \omega_{1,i} = \omega_{i,1} = 2.1$ and $\omega_{i,j} = 0$. otherwise.
9. *topology 9*: a circular model.
   - $\mathbf{\Omega}_1$: $\omega_{ii} = 2, \omega_{i,i-1} = \omega_{i-1,i} = 1$ and $\omega_{1,p} = \omega_{p,1} = 0.45$.
   - $\mathbf{\Omega}_2$: $\omega_{ii} = 4, \omega_{i,i-1} = \omega_{i-1,i} = 2$ and $\omega_{1,p} = \omega_{p,1} = 0.95$.

Loss functions. To assess the performance of DN matrix estimation for the D-net and B-net, five loss functions are considered below

- Frobenius norm: $\|\hat{\mathbf{\Delta}} - \mathbf{\Delta}\|_F$, where $\|\boldsymbol{A}\|_F = \sqrt{\text{trace}(\boldsymbol{A}\boldsymbol{A}^\top)}$. Referred to as F norm.
- L_1 eigenvalue loss: $\sum_{i=1}^{p}(\hat{\gamma}_i - \gamma_i)^2/p$. Referred to as EL1.
- L_2 eigenvalue loss: $\sum_{i=1}^{p}(\hat{\gamma}_i - \gamma_i)^2/p$. Referred to as EL2.
- L_1 loss on the largest eigenvalue: $|\hat{\gamma}_{\max} - \gamma_{\max}|$. Referred to as MAXEL1.
- L_1 loss on the smallest eigenvalue: $|\hat{\gamma}_{\min} - \gamma_{\min}|$. Referred to as MINEL1.

Notice that some loss functions utilise the true DN matrix and its estimates, while others utilise the eigenvalues and their respective estimates. The eigenvalue based loss functions are designed to investigate the extremes of the eigenvalue spectrum. In particular, the MAXEL1 loss function highlights which estimator is favourable in a principal component setting, (Banerjee et al., 2013).

Structure learning performance functions. The performance of the network recovery for the D-net and B-net are assessed using

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

- True positive (TP): an edge is correctly predicted.
- False positive (FP): an edge is incorrectly predicted.
- True negative (TN): the absence of an edge is correctly predicted.
- False negative (FN): the absence of an edge is incorrectly predicted.
- Sensitivity: $\frac{\text{TP}}{\text{TP+FN}}$.
- Specificity: $\frac{\text{TN}}{\text{TN+FP}}$.
- Precision: $\frac{\text{TP}}{\text{TP+FP}}$.
- Recall: $\frac{\text{TP}}{\text{2TP+FN}}$.
- F1 score: $\frac{\text{2TP}}{\text{2TP+FN+FP}}$.
- False negative rate (FNR): $\frac{\text{FP}}{\text{FP+TN}}$.
- Matthews Correlation Coefficient (MCC): $\frac{\text{TP}\times\text{TN}-\text{FP}\times\text{FN}}{\sqrt{(\text{TP+FP})(\text{TP+FN})(\text{TN+FP})(\text{TN+FN})}}$.

A description on the scope of these metrics is provided in Section 3.5 of Chapter 3. Further insights on the performance metrics are discussed in (Iwendi et al., 2020b).

Simulation results. Table 5.1 reports the median of F norm, EL1, EL2, MAXEL1 and MINEL1 for $p = 10, 30, 100$ in structures $1-9$ based on 40 replications. For the B-net, the loss values are calculated using the mean of the posterior distributions for $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$. For each scenario, the best performing measure is boldfaced. It is worth mentioning that a departure from the formatting of the synthetic analysis results in Chapters 3 and 4 is being used here to maintain consistency of the original work published in (Smith et al., 2022b).

A couple of observations are worth noting from Table 5.1 and Table 5.2. First, the D-net estimator performs better with the AR(1) structure. Second, the B-net estimator is generally the top performer in the remaining topologies. Third, the standard errors for both DN estimation techniques remain relatively consistent throughout the dimension spectrum considered, noting that the D-net estimator yields, in general, better results. This may be due to the fact that the best performing tuning parameter in the solution path leads to highly sparse estimates. The B-net estimation procedure inherits the utilisation of multiple penalty parameters in the precision matrix estimation, leading to robust estimation of the precision matrices.

To assess the performance on graphical structure determination, the specificity, sensitivity, FNR, F1 score and the MCCs are used. Values of specificity, sensitivity, F1 score and MCC closer to one, imply better classification performance. The closer the values of FNR are to zero, the better. The sparsity for the B-net estimator is determined by the thresholding rule in (5.13). Similarly, the best performing tuning parameter in the solution path of the D-net algorithm determines the sparsity of the estimator. The median performance scores, based on 40 repetitions, for each graphical structure are presented in Table 5.3. The main diagonals of the adjacency matrices are not included in the scoring.

The B-net estimator generally outperforms the D-net estimator across all models and all dimensions according to the MCC, F1 score, sensitivity and FNR, with the exception of the star and scale-free topologies - this is unsurprising since the D-net estimates a completely sparse network. Figure 5.4 display the true and inferred undirected DN graphs for both the B-net and D-net estimators for $p = 10$.

© University of Pretoria

Table 5.1 *Synthetic study median loss results.* Summary of F norm, EL1, EL2, MAXEL1 and MINEL1 for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model. The median loss values reported here are based on 40 replications for both the B-net and D-net estimators. The best performing values are boldfaced.

	AR(1)		**AR(2)**		**S80**		**S40**		**SF**		**Band**		**Cluster**		**Star**		**Circle**	
	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net
									p=10									
F norm	0.91	**0.64**	**1.41**	2.09	**2.21**	3.8	**6.09**	7.45	**2.14**	2.22	**1.39**	1.91	**1.05**	2.48	8.73	**8.49**	**3.87**	4.3
EL1	0.14	**0.1**	**0.24**	0.53	**0.23**	0.61	**0.63**	1.12	**0.11**	0.48	**0.22**	0.43	**0.25**	0.58	**1.19**	1.2	**0.42**	1.23
EL2	0.03	**0.03**	**0.09**	0.35	**0.1**	0.47	**0.47**	1.73	**0.02**	0.49	**0.07**	0.21	**0.1**	0.52	**6.02**	7.2	**0.27**	1.83
MAXEL1	**0.17**	0.51	**0.49**	1.07	**0.2**	0.72	**0.7**	1.57	**0.15**	1.37	**0.44**	0.55	**0.64**	1.42	**5.45**	6	**0.58**	1.91
MINEL1	0.28	**0.06**	**0.09**	0.45	**0.16**	0.67	**0.9**	1.09	**0.15**	1.37	**0.44**	0.49	0.36	**0.22**	**5.52**	6	**0.6**	1.91
									p=30									
F norm	1.98	**1.35**	**1.93**	3.82	**13.11**	17.03	**26.02**	32	3.77	**2.87**	**2.21**	3.67	8.39	**8.2**	15.26	**15.23**	**4.96**	7.65
EL1	0.16	**0.12**	**0.17**	0.58	**0.77**	2.4	**1.75**	4.67	**0.22**	0.31	**0.22**	0.59	**0.72**	0.74	**0.71**	0.72	**0.15**	1.25
EL2	**0.04**	0.06	**0.04**	0.43	**0.75**	7.36	**3.93**	28.49	**0.1**	0.27	**0.07**	0.42	**2.1**	2.21	**7.29**	7.73	**0.04**	1.92
MAXEL1	**0.56**	1.07	**0.22**	1.2	**1.04**	3.41	**2.84**	8.34	**1.05**	1.62	**0.49**	0.81	**5.41**	5.55	**10.45**	10.77	**0.64**	1.93
MINEL1	0.38	**0.14**	**0.16**	0.53	**1.14**	3.4	**3.03**	8.23	**0.85**	1.62	**0.49**	0.81	**0.19**	0.35	**10.45**	10.77	**0.64**	1.93
									p=100									
F norm	3.36	**2.62**	**3.1**	7.07	**58.3**	62.13	**103.75**	108.41	5.56	**3.88**	**3.87**	7.04	28.2	**28**	**28.17**	28.19	**9.17**	14.09
EL1	0.14	**0.13**	**0.15**	0.63	**3.37**	5.17	**5.04**	9.08	**0.11**	0.23	**0.27**	0.63	**0.78**	0.78	**0.4**	0.4	**0.18**	1.27
EL2	**0.03**	0.07	**0.03**	0.5	**14.36**	36.49	**33.86**	114.22	**0.03**	0.15	**0.1**	0.49	**7.69**	7.83	**7.78**	7.83	**0.06**	1.98
MAXEL1	**0.46**	1.29	**0.07**	1.35	**6.22**	9.97	**10.41**	18.07	**0.1**	1.77	**0.53**	1	**19.4**	19.58	**19.72**	19.79	**0.83**	1.98
MINEL1	0.48	**0.17**	**0.31**	0.67	**6.18**	10.04	**9.86**	17.81	**0.1**	1.77	**0.53**	1	**0.2**	0.38	**19.72**	19.79	**0.83**	1.98

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

© University of Pretoria

Table 5.2 *Synthetic study standard error loss results.* Summary of F norm, EL1, EL2, MAXEL1 and MINEL1 for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model. The median standard errors reported here are based on 40 replications for both the B-net and D-net estimators. The best performing values are boldfaced.

	AR(1)		**AR(2)**		**S80**		**S40**		**SF**		**Band**		**Cluster**		**Star**		**Circle**	
	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net
									p=10									
F norm	0.15	**0.03**	0.15	**0.04**	**0.78**	1.02	**1.07**	2.11	0.09	**0.01**	0.37	**0.15**	0.37	**0.05**	0.42	**0.06**	0.89	**0.01**
EL1	0.02	**0.01**	0.07	**0.04**	**0.12**	0.24	**0.19**	0.77	0.02	**0.01**	**0.09**	0.1	0.08	**0.03**	0.06	**0.02**	0.14	**0.03**
EL2	**0.01**	**0.01**	**0.05**	**0.05**	**0.11**	0.41	**0.39**	3.51	0.01	**0.01**	**0.06**	0.1	0.09	**0.04**	0.8	**0.29**	0.17	**0.1**
MAXEL1	0.11	**0.01**	0.25	**0.08**	**0.27**	0.5	**0.57**	1.28	0.08	**0.01**	**0.17**	0.18	0.3	**0.08**	0.4	**0.12**	0.48	**0.1**
MINEL1	0.13	**0.04**	0.1	**0.08**	**0.26**	0.57	**0.64**	1.46	0.08	**0.01**	**0.17**	0.22	0.12	**0.09**	0.38	**0.12**	0.5	**0.1**
									p=30									
F norm	0.2	**0.01**	0.14	**0.01**	**1.03**	1.23	**1.07**	1.23	0.05	**0.01**	0.27	**0.2**	0.04	**0.01**	0.06	**0.01**	1.07	**0.01**
EL1	0.02	**0.01**	0.03	**0.02**	**0.2**	0.25	**0.18**	0.3	**0.01**	**0.01**	**0.04**	**0.04**	**0.01**	**0.01**	**0.01**	**0.01**	0.05	**0.01**
EL2	**0.01**	**0.01**	**0.01**	0.03	**0.39**	1.58	**0.86**	4.6	**0.01**	**0.01**	**0.03**	0.07	0.03	0.02	0.15	**0.05**	0.05	**0.03**
MAXEL1	0.15	**0.02**	0.12	**0.06**	**0.61**	1.31	**0.76**	1.58	0.06	**0.01**	**0.1**	0.11	0.04	**0.03**	0.11	**0.04**	0.28	**0.04**
MINEL1	0.16	**0.02**	0.07	**0.05**	**0.61**	1.37	**0.83**	1.6	0.03	**0.01**	**0.11**	0.12	0.05	**0.03**	0.11	**0.04**	0.28	**0.04**
									p=100									
F norm	0.2	**0.01**	0.05	**0.01**	1.17	**1.1**	1.85	**1.71**	**0.01**	**0.01**	0.02	**0.01**	0.02	**0.01**	**0.01**	0.05	1.31	**0.01**
EL1	**0.01**	**0.01**	**0.01**	**0.01**	0.17	**0.11**	0.22	**0.16**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	0.04	**0.01**
EL2	**0.01**	**0.01**	**0.01**	**0.01**	**1.45**	1.56	**2.98**	3.84	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.06**	0.08	**0.01**	**0.01**
MAXEL1	0.19	**0.01**	0.07	**0.01**	**0.75**	0.89	**1.08**	1.32	0.03	**0.01**	**0.03**	**0.03**	**0.01**	**0.01**	**0.07**	0.1	0.14	**0.01**
MINEL1	0.12	**0.01**	0.05	**0.01**	**0.6**	0.89	**0.89**	1.52	0.03	**0.01**	**0.03**	**0.03**	**0.01**	**0.01**	**0.07**	0.1	0.14	**0.01**

© University of Pretoria

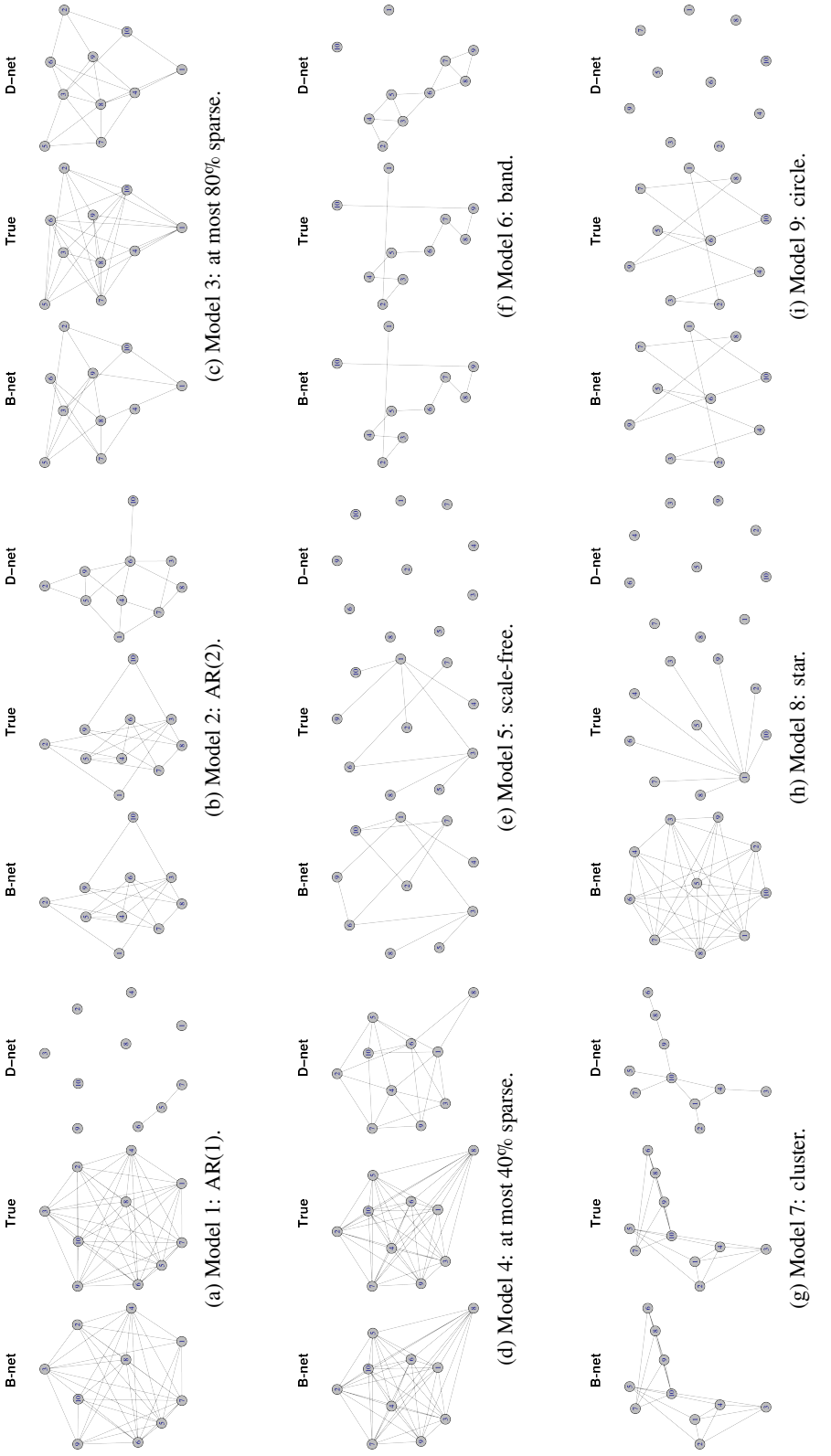
UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 5.3 *Synthetic study median performance results.* Summary of SE, SP, MCC, F1 and FNR for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model. The median performance scores reported here are based on 40 replications for both the B-net and D-net estimators. The best performing values are boldfaced.

	AR(1)		**AR(2)**		**S80**		**S40**		**SF**		**Band**		**Cluster**		**Star**		**Circle**	
	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net	B-net	D-net
p=10																		
SE	**0.72**	0.11	**0.82**	0.21	**0.6**	0.53	**0.65**	0.36	**0.89**	0	**0.89**	0.44	**1**	0.25	**0.22**	0.11	**0.9**	0
SP	**1**	**1**	**0.97**	0.7	**0.89**	0.82	0.87	0.94	0.9	**1**	**0.95**	0.93	**1**	0.87	**0.8**	1	0.95	**0.98**
MCC	**0.45**	0.11	**0.76**	-0.07	**0.54**	0.39	**0.43**	0.24	**0.71**	0	**0.83**	0.41	**1**	0.14	**0.11**	0.3	**0.88**	-0.07
F1	**0.84**	0.2	**0.84**	0.24	**0.67**	0.56	**0.77**	0.51	**0.76**	0	**0.86**	0.4	**1**	0.34	**0.25**	0.2	**0.9**	0
FNR	**0.28**	0.89	**0.18**	0.79	**0.4**	0.47	**0.35**	0.64	0.11	**0**	**0.11**	0.56	**0**	0.75	**0.78**	0.89	**0.1**	1
p=30																		
SE	**0.31**	0.01	**0.75**	0.04	**0.23**	0.08	**0.37**	0.03	**0.79**	0	**0.9**	0.24	**0.89**	0.02	**0.03**	0.02	**0.97**	0
SP	0.97	**1**	**0.99**	0.93	0.98	**0.99**	0.82	**1**	**0.92**	1	**1**	**1**	0.99	0.98	**1**	**1**	**1**	0.99
MCC	**0.17**	0.02	**0.8**	-0.04	**0.35**	0.16	**0.16**	0.05	**0.52**	0	**0.92**	0.46	**0.89**	-0.02	0.06	**0.08**	**0.95**	-0.03
F1	**0.48**	0.01	**0.82**	0.04	**0.36**	0.14	**0.53**	0.06	**0.53**	0	**0.93**	0.38	**0.93**	0.04	**0.06**	0.03	**0.95**	0
FNR	**0.69**	0.99	**0.25**	0.96	**0.77**	0.92	**0.63**	0.97	**0.21**	0	**0.1**	0.76	**0.11**	0.98	**0.97**	0.98	**0.03**	1
p=100																		
SE	**0.21**	0	**0.73**	0	**0.04**	0.01	**0.15**	0	**0.49**	0	**0.95**	0.05	**0.79**	0	0.02	**0.12**	**0.99**	0
SP	**0.99**	1	**1**	**1**	**1**	**1**	0.92	**1**	0.98	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
MCC	**0.33**	0.01	**0.82**	0	**0.14**	0.05	**0.07**	0.01	**0.39**	0	**0.95**	0.22	**0.81**	0	0.08	**0.35**	**0.98**	-0.01
F1	**0.34**	0	**0.82**	0	**0.08**	0.01	**0.25**	0	**0.4**	0	**0.95**	0.1	**0.88**	0	0.04	**0.22**	**0.98**	0
FNR	**0.79**	1	**0.27**	1	**0.96**	0.99	**0.85**	1	0.51	**0**	**0.05**	0.95	**0.21**	1	0.98	**0.88**	**0.01**	1

© University of Pretoria



(a) Model 1: AR(1). (b) Model 2: AR(2). (c) Model 3: at most 80% sparse.

(d) Model 4: at most 40% sparse. (e) Model 5: scale-free. (f) Model 6: band.

(g) Model 7: cluster. (h) Model 8: star. (i) Model 9: circle.

Fig. 5.4 *Estimated DN graphical networks.* Comparison of the true DN, B-net and D-net graphical structure estimates for an AR(1), AR(2), sparse random, scale-free, band, cluster, star and circle graphical model and $p = 10$.

© University of Pretoria

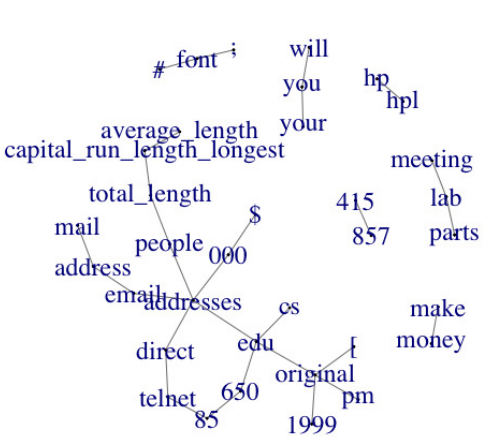
5.4 Real data analysis

This section focuses on applying the Bayesian DN estimator, B-net, as well as the iterative shrinkage-thresholding estimator, D-net, to the spambase dataset to investigate changes in DN structure between spam and non-spam data. In addition, the B-net estimator is applied to South African COVID-19 data, obtained from three data sources available from the South African National Institute for Communicable Diseases, Our World in Data and Media Hack Collective to investigate the change in DN structure between various phases of the pandemic.

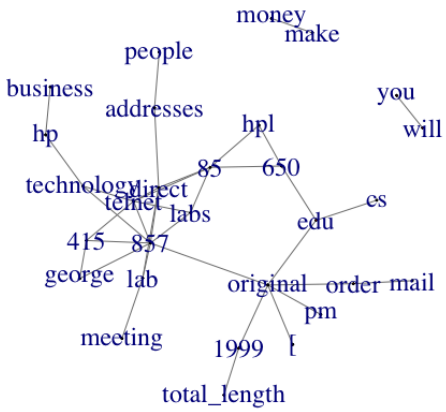
5.4.1 Spam data

The objective here is to compare the B-net and D-net graphical model estimates of the spam and non-spam emails. The dataset consists of 1813 spam emails and 2788 non-spam emails. The attributes include, amongst others, the average length of uninterrupted sequences of capital letters; total number of capital letters in the e-mail; an indicator denoting whether the e-mail was considered spam or not, in this study.

Following the approach of (Tang et al., 2020), the data is standardised using a non-paranormal transformation in order to satisfy the Gaussian assumption. The B-net estimates are based on 10 000 iterations of the Monte Carlo sampler after 5000 burn-in iterations. Figure 5.5 illustrates the difference between the B-net and D-net estimates. Both estimators indicate the presence of several common hub features namely, 'edu', 'original', 'direct', 'lab', 'telnet' and 'addresses'. It is clear from both panes that the state of the covariance matrix structure between the spam and non-spam emails may very well be different. Furthermore, given that Hewlett-Packard Labs donated the data, words such as 'telnet' and 'hp' appear more often in the non-spam emails and can be used to distinguish between spam and non-spam emails.



(a) The Bayesian DN for the spam emails dataset.



(b) The iterative-shrinkage DN for the spam emails dataset.

Fig. 5.5 *A comparison of the D-net and B-net DN estimates for the spam emails dataset.*

© University of Pretoria

5.4.2 South African COVID-19 data

A momentary departure from the prevailing use of the present tense in the writing style is necessary in this application, given that the coronavirus pandemic has largely run its course at the time of writing this dissertation. The 2019 novel coronavirus (COVID-19) affected more than 180 countries around the world, including South Africa. The current body of knowledge boasts a wealth of statistical literature that aims at empowering researchers to study and alleviate the impact or better prepare for similar diseases, see for example (Iwendi et al., 2020a). Understanding the interaction of key metrics and attributes between various phases, cycles or waves of the pandemic may prove to be invaluable in strategic planning and prevention. The goal here, is to use the Bayesian DN, B-net, to illustrate that the interactivity of key daily metrics between suspected homogeneous and heterogeneous phases within the pandemic life cycle was ever changing. In particular, the B-net is used to model the interactivity of daily metrics between the first two peaks or waves; the first wave and the following plateau and finally the difference between the first and second post wave plateaus. The data consists of 446 observations from the 7th of February 2020 to the 27th of April 2021. The daily metrics include, deaths; performed tests; positive test rate; active cases; tests per active case; recoveries; hospital admissions; hospital discharges; ICU admissions and the number of ventilated patients. It should be noted that no sensitive patient information is used, however, the interested reader is referred to (Iwendi et al., 2020c) for a detailed treatment and framework for dealing with and sanitising medical data containing sensitive patient information. Due to the irregularities in data capturing and publishing, a seven day moving average is applied across all daily metrics. The data is standardised using a non-paranormal transformation in order to satisfy the Gaussian assumption. The B-net is applied to the data using 10 000 iterations of the Monte Carlo sampler after 5000 burn-in iterations.

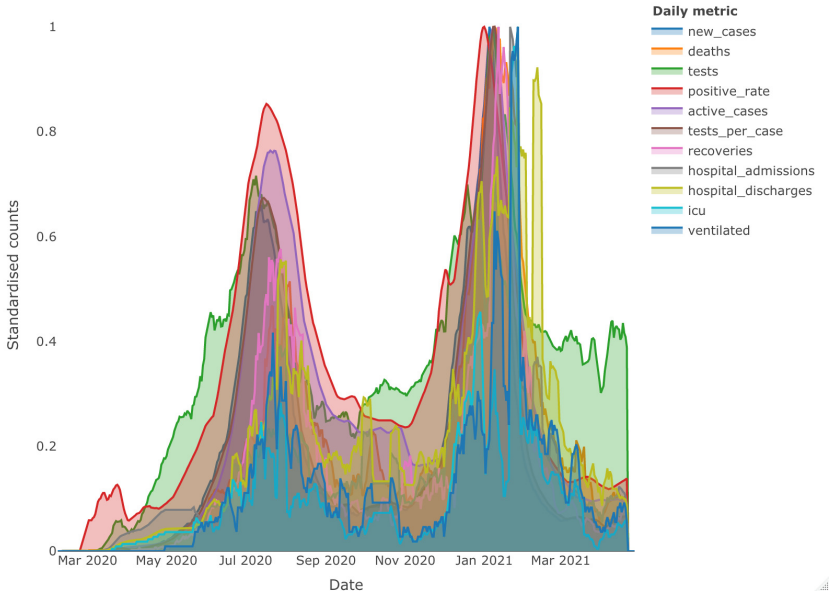


Fig. 5.6 *South African COVID-19 daily metrics over time.* 7-day moving average filled area line plots with standardised counts for daily new cases; deaths; tests; positive test rate; active cases; tests per active case; recoveries; hospital admissions; hospital discharges; ICU admissions and ventilated patients.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Figure 5.6 highlights the temporal nature of the pandemic between suspected homogeneous and heterogeneous phases. In other words, comparing the cyclical behaviour of individual daily metrics may seem clearly distinctive over time; a peak or wave is always followed by a plateau. Furthermore, extrapolation of the temporal behaviour of individual daily metrics may incorrectly allude to distinct multimodality of multiple daily metrics. Upon observing multiple metrics simultaneously, the crisp group-wise multimodality diminishes rather rapidly. Figure 5.7 illustrates the higher proportions of hub features present in the DNs. Interestingly, the Bayesian DN provides insight to the change in interaction between daily metrics and between perceived homogeneous pandemic phases. In other words, comparisons between the two peaks and two post-peak plateaus. This change in behaviour could be as a result of the change in population adherence to public sanitation awareness; weather conditions; virus mutations or complacency over time.

© University of Pretoria

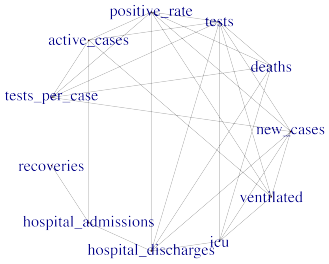
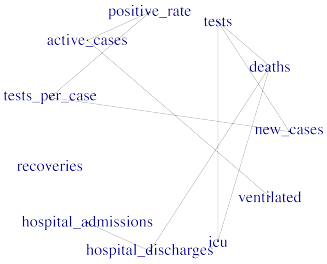
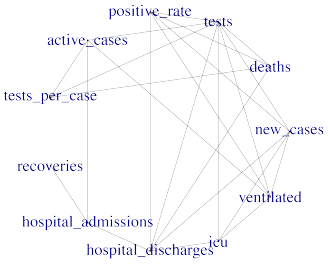
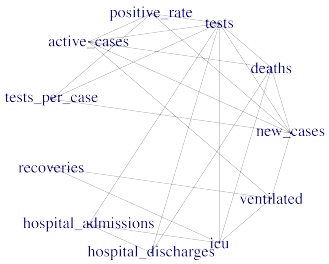
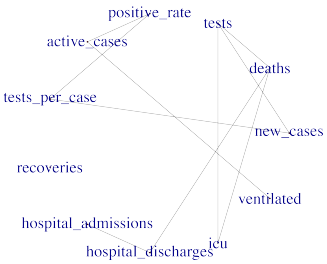
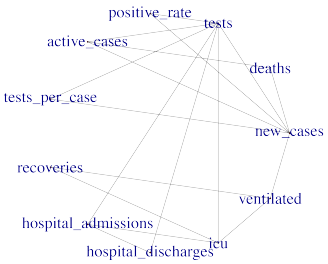
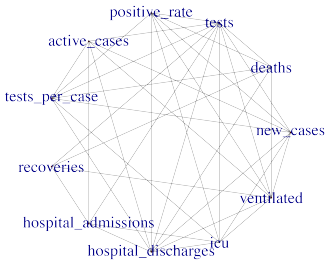
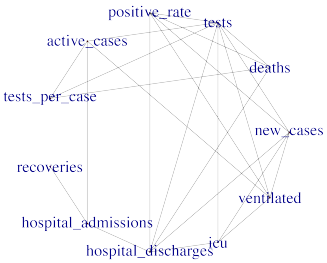
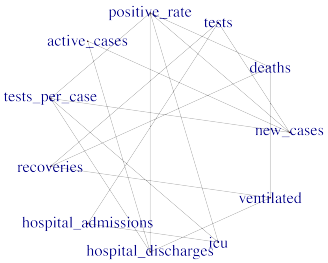
Bayesian DN	First precision	Second precision
		
Bayesian DN for first wave and first post wave plateau.	BAGL graphical model of the first wave.	BAGL graphical model of the first post wave plateau.
		
Bayesian DN for first and second wave.	BAGL graphical model of the first wave.	BAGL graphical model of the second wave.
		
Bayesian DN for first and second post wave plateaus.	BAGL graphical model of the first post wave plateau.	BAGL graphical model of the second post wave plateau.

Fig. 5.7 *Bayesian DN estimates of South African COVID-19 data.* The Bayesian DN and corresponding BAGL graphical models between the first two waves; the first wave and the following plateau and finally the difference between the first and second post wave plateaus. The $p-values$ from the Box's M-test for homogeneity of covariance matrices between the contributing precision matrices were all less than 0.001 (Box, 1949).

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

5.5 Discussion

At the time of writing, the Bayesian differential network estimator is the first of its kind. The methodology separately estimates the components (precision matrices) of the differential network using the Bayesian adaptive graphical lasso (Wang, 2012). Structure learning is achieved through a simple, model agnostic Wishart thresholding technique. In comparison with the state-of-the-art iterative shrinkage-thresholding approach, the Bayesian differential network offers MCMC outputs that allow the user to gain deeper insight and inference in the estimation procedure - see for example Chapters 3 and 4. The numerical accuracy of the Bayesian differential network is, in general, superior to that of the iterative shrinkage-thresholding estimator. The latter heavily prioritises sparsity. From a practical application perspective, the Bayesian differential network also provides key insights to changes in the interactive behaviour of real data metrics ranging from filtering spam emails to COVID-19 life cycles. For the cases where $n > p$ the Bayesian adaptive graphical ridge-type and naïve Bayesian adaptive graphical elastic net can be used in place of the Bayesian adaptive graphical lasso used here. One particular topic of interest is currently under investigation, albeit a difficult one: research on simultaneous Bayesian estimation of both $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ in the construction of the differential network.

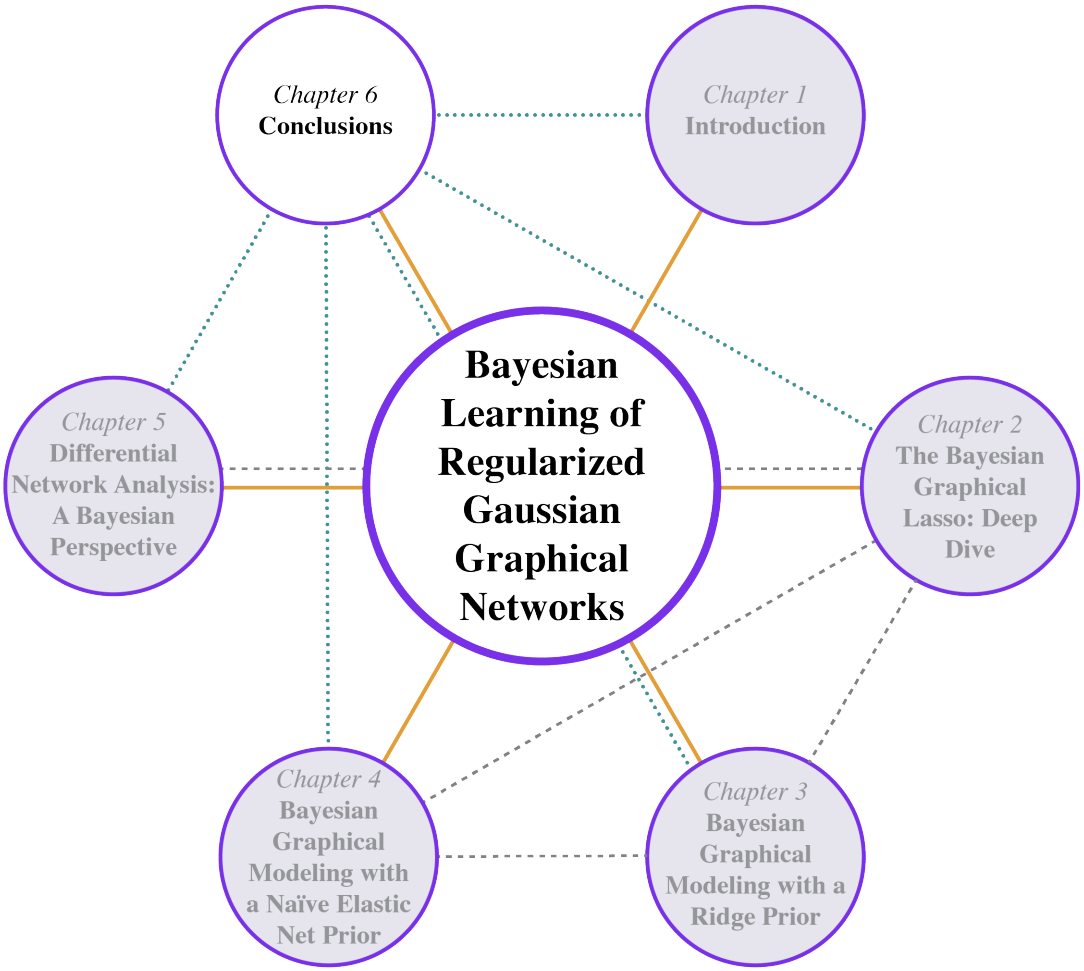
© University of Pretoria

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chapter 6

Conclusions



© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

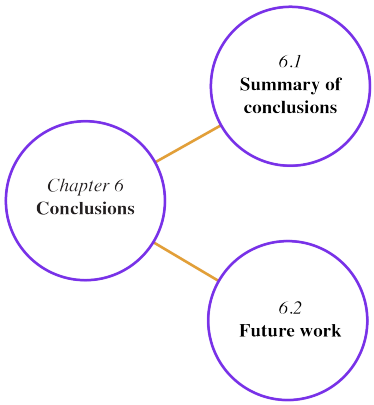


Fig. 6.1 Chapter 6 visual outline.

The conclusions of the work presented in this dissertation are summarised in this chapter. Figure 6.1 provides a visual outline of this chapter.

6.1 Summary of conclusions

In our increasingly complex and interconnected multivariate data age, the ability to infer the presence and magnitude of associational dependencies among variables becomes ever more essential. Gaussian graphical models offer *an* approach to modelling these associations by encoding them in the off-diagonal elements of the precision matrix. Under the Bayesian lens, there has been a rapid increase of research dedicated towards this approach. The parameters of interest within the Bayesian paradigm of GGM estimation are the graph G and precision matrix. The advantage of the Bayesian approach is the richness of information provided from drawing samples from the posterior distribution. That being said, Bayesian methods are known to struggle with the computational costs of scaling to higher dimensional variable problems. Bayesian methodologies focusing on the space of precision matrices aim to address this issue. This dissertation contributes to the tools available to Statisticians working at the front line of multivariate analysis by enriching the field of Bayesian shrinkage methodologies that focus on the space of precision matrices with varying sparsity requirements.

The mathematical premises of the Bayesian graphical lasso are unpacked. Here, emphasis is placed on the formulation of the block Gibbs sampler and the extension of the original prior to allow for adaptive shrinkage of the off diagonal elements of the precision matrix. These building blocks serve as the foundation for the priors, samplers and applications considered subsequently.

Inspired by the motivation for frequentist graphical ridge in (Wieringen and Peeters, 2016), the novel Bayesian graphical ridge prior is proposed. The prior is constructed as the product of independent Gaussian pdfs for

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

the off-diagonal elements of the precision matrix and the product of truncated Gaussians pdfs for the diagonal elements. It is shown that the posterior mode resulting from this prior is equivalent to the objective function of the frequentist graphical ridge, in particular their alternative type II formulation with a null target matrix. A full block Gibbs sampler for implementing this estimator is provided. The Bayesian graphical ridge prior is extended to the Bayesian adaptive graphical ridge hierarchical prior. The hierarchical prior makes use of inverse gamma pdfs for the shrinkage parameters of each off-diagonal element. The Bayesian adaptive graphical ridge displays commendable performance when compared to the frequentist alternative type I and II estimators and out performs the Bayesian graphical lasso for denser topologies when $n \leq p$, and when $p > n$ across all topologies considered.

To enable higher flexibility in data analysis where the sparsity requirements are not known apriori, the novel naïve Bayesian elastic net prior is proposed. The prior is constructed by combining the components of the Bayesian graphical ridge and lasso priors. The posterior mode resulting from this prior is equivalent to the objective function of the graphical lasso, but includes both the ℓ_1 and ℓ_2 penalties on the precision matrix. A full block Gibbs sampler for implementing this estimator is provided. Similarly to the Bayesian adaptive graphical lasso and ridge hierarchical priors, an adaptive extension of the naïve Bayesian graphical elastic net prior is proposed. That is, the Bayesian framework here facilitates the usage of hyperpriors on the ℓ_1 and ℓ_2 shrinkage parameters. In particular, two formulations are considered; the first employs the hyperpriors used by the Bayesian adaptive graphical lasso and ridge, however, the caveat here is the presence of four hyperparameters. The second formulation considers, instead exponential hyperpriors on the ℓ_1 shrinkage parameter and on the inverse of the ℓ_2 shrinkage parameter - reducing the required hyperparameters. Synthetic and practical data analyses demonstrate the ability of the naïve Bayesian adaptive graphical elastic net to offer a balanced performance in the presence of varying sparsity and p/n ratios when compared to its ridge and lasso counterparts.

Finally, a Bayesian differential network estimator is proposed for scenarios where networks may undergo changes over time or where they may differ between states. The methodology separately estimates the components (precision matrices) of the differential network using the Bayesian adaptive graphical lasso, however, the Bayesian (adaptive) graphical ridge or naïve Bayesian (adaptive) graphical elastic net can be used instead. Structure learning is achieved through a simple, model agnostic Wishart thresholding technique. The numerical accuracy of the Bayesian differential network is, in general, superior to that of the frequentist iterative shrinkage-thresholding estimator, which heavily prioritises sparsity. Differential network analysis holds great promise for a wide range of biological applications ((Ideker and Krogan, 2012), (Hsiao et al., 2016), (Troy et al., 2016), (Sas et al., 2018), (Singh et al., 2018)) and the ongoing need for statistical methods to identify network differences underscores their importance in this field. The potential of Bayesian methodologies lies within the ability to incorporate prior information, for example, the magnitude and presence of known cancer-related genes in subtypes of cancer (Shojaie, 2020). An advantage of the proposed framework is that there is no requirement nor prior assumption required on the similarity of the structure or edge association strengths of the components of the DN.

In summary: Bayesian priors for a ridge and naïve elastic net, rooted in absolute continuous priors and arbitrary sparsity patterns, are proposed for the estimation of precision matrices in the setting of Gaussian graphical models. The proposed techniques compare favourably to existing state-of-the-art frequentist and Bayesian methodologies in simulations and practical data analyses with moderate dimensions. The most prominent benefits are in situations where accurate estimation of the precision matrix is required and where sparsity is not always desired or where a balanced treatment thereof is. Finally, differential network estimation using Bayesian precision matrix estimation techniques is presented. Simulation results illustrate that the Bayesian differential network estimator performs favourably and underscores the advantages of employing the Bayesian techniques assessed in this dissertation.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

6.2 Future work

A challenge in GGM simulation studies is designing a representative simulation that balances various factors, such as the sparsity of precision matrices, the observation-to-variable ratio, and the size of precision matrix elements. Evaluating a proposed technique requires considering numerous combinations, which can be quite complex and general guidance in this regard is not available in literature. A more elegant approach would be to derive and study theoretical properties of proposed estimators, that could assist with gauging the expected performance, limitations and asymptotics of these estimators. Theoretical insights, in this regard, were not explored for the novel priors presented in Chapters 3 and 4, but it would be of interest to do so. For example the theoretical bounds of the Kullback-Liebler divergence and element-wise bias for varying observation-to-variable ratios, see (Li et al., 2019) for an investigation on the Bayesian graphical horseshoe.

The scale parameter of the inverse gamma pdf in Algorithm 2 may misbehave if exploited. In particular if this parameter explodes then the resulting distribution will be narrower and more concentrated which can be considered as a bonus to the sampling scheme. The opposite may have undesirable effects. One potential remedy includes considering the kappa-mu inverse gamma fading model instead.

With regards to graphical structure learning, the heuristic approaches in this dissertation did not consider a formal Bayesian treatment - which could address the caveat of the Bayesian techniques focused purely on precision matrix estimation Vogels et al. (2023). Moreover, credible interval-based structure learning requires further investigation in two directions. The first being the coverage properties of marginal credible intervals for the priors proposed in Chapter 3 and 4 to understand the effect on the false positives since wider intervals lead to sparser solutions. Next, to determine whether the positive definite constraint on the precision matrix is maintained once sparsity is enforced. It would also be interesting to evaluate the performance of the geometric credible interval approach in (Khondker et al., 2013); the penalised credible interval approach in (Kundu et al., 2019) and the multiple hypothesis testing based approach in (Chandra et al., 2022).

Concerning the formulation of the naïve Bayesian graphical elastic net prior in Chapter 4, an alternative prior formulation can be investigated which avoids the need for two shrinkage parameters and corresponds more closely to the frequentist graphical elastic net objective function in (Kovács et al., 2021). That is a Bayesian treatment of the 'proper' elastic net penalty with a hyperparameter α controlling how much weight is assigned to each of the ℓ_1 and ℓ_2 penalties.

The Bayesian differential network estimator in Chapter 5 separately estimates the components of the differential network. This formulation, albeit not explored, can assist in testing the global null hypothesis of no difference between two GGMs (Shojaie, 2020) by using the richness of the posterior distribution for each component within the Bayesian hypothesis testing techniques in (Williams et al., 2018). A challenge with this approach, however, is that a symmetric network estimate is not guaranteed. The development of a Bayesian estimator which estimates the differential network directly may address this issue, in addition to quantifying uncertainty when testing for differences in network edges. An obvious challenge here is the prior formulation for the difference of two Wishart random variables. In addition to this, future simulation studies can include scenarios where the structure of the components of the DN need not be equal. Lastly, differential network estimation techniques where the components of the differential network are estimated separately largely assume that the observations are stochastically independent. To this end, future work may consider a extending the two-layer Gaussian graphical model estimation technique in (Xie et al., 2016) to the differential network domain and then investigating a Bayesian treatment thereof. The technique of interest addresses the estimation of multiple dependent GGMs by means of two-layer estimation approach. The first layer addresses systemic variation between all variables of interest, whereas the second layer focusses on category-specific variation.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Alas, the resounding conclusion is the potential utility of the Bayesian methodological extensions derived in this dissertation, enabling higher flexibility for multivariate data analysis.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

References

Andrews, D. and Mallows, C. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(1):99–102.

Atay-Kayis, A. and Massam, H. (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika*, 92(2):317–335.

Banerjee, O., Ghaoui, L. E., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516.

Banerjee, S. and Ghosal, S. (2015). Bayesian structure learning in graphical models. *Journal of Multivariate Analysis*, 136:147–162.

Banerjee, S., Monni, S., and Wells, M. (2013). A regularized profile likelihood approach to covariance matrix estimation. *Journal of Statistical Planning and Inference*, 179:36–59.

Baz, J., Alonso, P., Peña, J., and Pérez-Fernández, R. (2023). Gaussian Markov random fields and totally positive matrices. *Journal of Computational and Applied Mathematics*, 430:115098.

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202.

Bhadra, A., Datta, J., Polson, N., and Willard, B. (2021). The horseshoe-like regularization for feature subset selection. *Sankhya B*, 83:185–214.

Box, G. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, 36(3/4):317–346.

Brown, P. and Griffin, J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188.

Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Carvalho, C., Polson, N., and Scott, J. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Carvalho, P. H. C. (2015). Decoupling shrinkage and selection in Bayesian linear models: a posterior summary perspective. *Journal of the American Statistical Association*, 110(509):435–448.

Chandra, N., Muller, P., and Sarkar, A. (2022). Bayesian scalable precision factor analysis for massive sparse Gaussian graphical models. *arXiv preprint arXiv:2107.11316*.

Chen, L. (2023). Estimation of graphical models: An overview of selected topics. *International Statistical Review*, https://doi.org/10.1111/insr.12552.

Cheng, Y. and Lenkoski, A. (2012). Hierarchical Gaussian graphical models: Beyond reversible jump. *Electronic Journal of Statistics*, 6:2309–2331.

Chhikara, R. and Folks, J. (1989). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. Marcel Dekker, New York.

Chicco, D. and Jurman, G. (2020). The advantages of the Matthews correlation coefficient (mcc) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(6).

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chiquet, J., Grandvalet, Y., and Ambroise, C. (2011). Inferring multiple graphical structures. *Statistics and Computing*, 21(4):537–553.

Dawid, A. and Lauritzen, S. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 21(3):1272–1317.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, 28(1):157–175.

Dobra, A., Lenkoski, A., and Rodriguez, A. (2011). Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *Journal of the American Statistical Association*, 106(496):1418–1433.

Drton, M., Massam, H., and Olkin, I. (2008). Moments of minors of Wishart matrices. *The Annals of Statistics*, 36:2261–2283.

Edwards, D. (2000). *Introduction to Graphical Modelling*. Springer, New York.

Fan, J., Feng, Y., and Wu, Y. (2009). Network exploration via the adaptive LASSO and SCAD penalties. *The Annals of Applied Statistics*, 3(2):521—541.

Fouskakis, D., Ntzoufras, I., and Draper, D. (2009). Bayesian variable selection using cost-adjusted BIC, with application to cost-effective measurement of quality of health care. *Annals Applied Statistics*, 3:663—-690.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7(4):457–472.

Golub, G. and Loan, C. V. (1996). *Matrix Computations*. JHU press, Baltimore.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.

Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika*, 98(1):1–15.

Haff, L. (1991). The variational form of certain Bayes estimators. *The Annals of Statistics*, 19(3):1163–1190.

Hsiao, T., Chiu, Y., Hsu, P., Lu, T., Lai, L., Tsai, M., Huang, T., Chuang, E., and Chen, Y. (2016). Differential network analysis reveals the genome-wide landscape of estrogen receptor modulation in hormonal cancers. *Scientific Reports*, 6:23035.

Ideker, T. and Krogan, N. (2012). Differential network biology. *Molecular Systems Biology*, 8(1):565.

Iwendi, C., Bashir, A., Peshkar, A., Sujatha, R., Chatterjee, J., Pasupuleti, S., Mishra, R., Pillai, S., and Jo, O. (2020a). Covid-19 patient health prediction using boosted random forest algorithm. *Frontiers in Public Health*, 8(357).

Iwendi, C., Khan, S., Anajemba, J., Mittal, M., Alenezi, M., and Alazab, M. (2020b). The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems. *Sensors*, 20(9):2559.

Iwendi, C., Moqurrab, S., Anjum, A., Khan, S., Mohan, S., and Srivastava, G. (2020c). N-sanitization: A semantic privacy-preserving framework for unstructured medical datasets. *Computer Communications*, 161:160–171.

Jiang, B., Wang, X., and C.Leng (2018). A direct approach for sparse quadratic discriminant analysis. *The Journal of Machine Learning Research*, 19(1):1098–1134.

Jones, B., Carvalho, C., Dobra, A., andC. Carter, C. H., and West, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science*, 20(4):388–400.

Keune, J., Ohlwein, C., and Hense, A. (2014). Multivariate probabilistic analysis and predictability of medium-range ensemble weather forecasts. *Monthly Weather Review*, 142(11):4074–4090.

Khondker, Z., Zhu, H., Chu, H., Lin, W., and Ibrahim, J. (2013). The Bayesian covariance lasso. *Statistics and its Interface*, 6(2):243—-259.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The Review of Economic Studies*, 65(3):361–393.

Kovács, S., Ruckstuhl, T., H.O., and Buhlmann, P. (2021). Graphical elastic net and target matrices: Fast algorithms and software for sparse precision matrix estimation.

Kuismin, M., Kemppainen, J., and Sillanpää, M. (2017). Precision matrix estimation with ROPE. *Journal of Computational and Graphical Statistics*, 26(3):682–694.

Kundu, S., Mallick, B., and Baladandayuthapan, V. (2019). Efficient Bayesian regularization for graphical model selection. *Bayesian Analysis*, 14(2):449–476.

Lauritzen, S. (1996). *Graphical Models*. Clarendon Press, Oxford.

Lenkoski, A. (2013). A direct sampler for G-Wishart variates. *Statistics*, 2(1):119–128.

Lenkoski, A. and Dobra, A. (2011). Computational aspects related to inference in Gaussian graphical models with the G-Wishart prior. *Journal of Computational and Graphical Statistics*, 20(1):140–157.

Leppä-Aho, J., Pensar, J., Roos, T., and Corander, J. (2017). Learning Gaussian graphical models with fractional marginal pseudo-likelihood. *International Journal of Approximate Reasoning*, 83:21–42.

Letac, G. and Massam, H. (2007). Wishart distributions for decomposable graphs. *The Annals of Statistics*, 35(3):1278–1323.

Li, Y., Craig, B., and Bhadra, A. (2019). The graphical horseshoe estimator for inverse covariance matrices. *Journal of Computational and Graphical Statistics*, 28(3):747–757.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462.

Mohammadi, A. and Wit, E. (2015). Bayesian structure learning in sparse Gaussian graphical models. *Bayesian Analysis*, 10(1):109–138.

Mohammadi, R., Massam, H., and Letac, G. (2023a). Accelerating Bayesian structure learning in sparse Gaussian graphical models. *Journal of the American Statistical Association*, 118(542):1345–1358.

Mohammadi, R., Schoonhoven, M., Vogels, L., and Birbil, I. (2023b). High-dimensional Bayesian structure learning in Gaussian graphical models using marginal pseudo-likelihood. *arXiv preprint arXiv:2307.00127*.

Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Ncw Jersey.

Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.

Peterson, C., Stingo, F., and Vannucci, M. (2015). Bayesian inference of multiple Gaussian graphical models. *Journal of the American Statistical Association*, 110(509):159–174.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reich, H. B. B. (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. *Journal of the American Statistical Association*, 107(500):1610–1624.

Roverato, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scandinavian Journal of Statistics*, 29(3):391–411.

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D., and Nolan, G. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529.

Sagar, K., Banerjee, S., Datta, J., and Bhadra, A. (2021). Precision matrix estimation under the horseshoe-like prior-penalty dual.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Sas, K., Lin, J., Rajendiran, T., Soni, T., Nair, V., Hinder, L., Jagadish, H., Gardner, T., F.C., S. A., and Brosius (2018). Shared and distinct lipid-lipid interactions in plasma and affected tissues in a diabetic mouse model. *Journal of Lipid Research*, 59(2):173–183.

Shin, Y., Remacle, F., Fan, R., Hwang, K., Wei, W., Ahmad, H., Levine, R., and Heath, J. (2011). Protein signaling networks from single cell fluctuations and information theory profiling. *Biophysical Journal*, 100(10):2378–2386.

Shojaie, A. (2020). Differential network analysis: A statistical perspective. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(2):e1508.

Shutta, K., Balzer, L., Scholtens, D., and Balasubramanian, R. (2023). Spiderlearner: An ensemble approach to Gaussian graphical model estimation. *Statistics in Medicine*, 42(13):2116–2133.

Silvester, J. (2000). Determinants of block matrices. *The Mathematical Gazette*, 84(501):460–467.

Singh, A., Ramsey, S., Filtz, T., and Kioussi, C. (2018). Differential gene regulatory networks in development and disease. *Cellular and Molecular Life Sciences*, 75:1013–1025.

Smith, J., Arashi, M., and Bekker, A. (2022a). A data driven Bayesian graphical ridge estimator. *arXiv preprint arXiv:2210.16290*.

Smith, J., Arashi, M., and Bekker, A. (2022b). Empowering differential networks using Bayesian analysis. *PLoS ONE*, 17(1):e0261193.

Smith, J., Bekker, A., and Arashi, M. (2023). A naıve Bayesian graphical elastic net: driving advances in differential network analysis. *arXiv preprint arXiv:2306.14199*.

Speed, T. and Kiiveri, H. T. (1986). Gaussian Markov distributions over finite graphs. *The Annals of Statistics*, 14(1):138–150.

Tang, Z., Yu, Z., and Wang, C. (2020). A fast iterative algorithm for high-dimensional differential network. *Computational Statistics*, 35(1):95–109.

Teng, J. (2019). Seer breast cancer data. IEEE Dataport.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Troy, N., Hollams, E., Holt, P., and Bosco, A. (2016). Differential gene network analysis for the identification of asthma-associated therapeutic targets in allergen-specific t-helper memory responses. *BMC Medical Genomics*, 9(1):9.

Uhler, C. (2017). Gaussian graphical models: An algebraic and geometric perspective. *arXiv preprint arXiv:1707.04345*.

van den Boom, W., Beskos, A., and Iorio, M. D. (2022a). The G-Wishart weighted proposal algorithm: Efficient posterior computation for Gaussian graphical models. *Journal of Computational and Graphical Statistics*, 31(4):1215–1224.

van den Boom, W., Jasra, A., Iorio, M. D., Beskos, A., and Eriksson, J. (2022b). Unbiased approximation of posteriors via coupled particle Markov chain Monte Carlo. *Statistics and Computing*, 32(3).

van Wieringen, W. (2019). The generalized ridge estimator of the inverse covariance matrix. *Journal of Computational and Graphical Statistics*, 28(4):932–942.

Vats, D. and Knudson, C. (2021). Revisiting the Gelman-Rubin diagnostic. *Statistical Science*, 36(4):518–529.

Vogels, L., Mohammadi, R., Schoonhoven, M., and Birbil, Ş. İ. (2023). Bayesian structure learning in undirected Gaussian graphical models: Literature review with empirical comparison. *arXiv preprint arXiv:2307.02603*.

Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis*, 7(4):867–886.

Wang, H. (2015). Scaling it up: Stochastic search structure learning in graphical models. *Bayesian Analysis*, 10(2):351–377.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Wang, H. and Li, S. (2012). Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electronic Journal of Statistics*, 6:168–198.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, 74(3):646–648.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. John Wiley and Sons, Chichester.

Wieringen, W. V. and Peeters, C. (2016). Ridge estimation of inverse covariance matrices from high-dimensional data. *Computational Statistics & Data Analysis*, 103:284–303.

Williams, D. (2021). Bayesian estimation for Gaussian graphical models: Structure learning, predictability, and network comparisons. *Multivariate Behavioral Research*, 56(2):336–352.

Williams, D., Piironen, J., Vehtari, A., and Rast, P. (2018). Bayesian estimation of Gaussian graphical models with predictive covariance selection. *arXiv preprint arXiv:1801.05725*.

Witten, D. and Tibshirani, R. (2009). Covariance-regularized regression and classification for high dimensional problems. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(3):615–636.

Xie, Y., Liu, Y., and Valdar, W. (2016). Joint estimation of multiple dependent Gaussian graphical models with applications to mouse genomics. *Biometrika*, 103(3):493–511.

Yang, R. and Berger, J. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, 22(3):1195–1211.

Yin, H., Liu, X., and Kong, X. (2020). Gaussian mixture graphical lasso with application to edge detection in brain networks. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 1430–1435.

Yin, J. and Li, H. (2011). A sparse conditional Gaussian graphical model for analysis of genetical genomics data. *The Annals of Applied Statistics*, 5(4):2630–2650.

Yuan, H., Xi, R., Chen, C., and Deng, M. (2017). Differential network analysis via lasso penalized d-trace loss. *Biometrika*, 104(4):755–770.

Yuan, M. (2008). Efficient computation of ℓ_1 regularized estimates in Gaussian graphical models. *Journal of Computational and Graphical Statistics*, 17(4):809–826.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35.

Zhao, S., Cai, T., and Li, H. (2014). Direct estimation of differential networks. *Biometrika*, 101(2):253–268.

Zhou, S., Rütimann, P., Xu, M., and Bühlmann, P. (2011). High-dimensional covariance estimation based on Gaussian graphical models. *The Journal of Machine Learning Research*, 12:2975–3026.

Zhu, Y. and Li, L. (2018). Multiple matrix Gaussian graphs estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 80(5):927–950.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320.

© University of Pretoria

© University of Pretoria

Appendix A

The Bayesian Graphical Lasso: Deep Dive

A.1 The Bayesian graphical lasso

A.1.1 Block Gibbs sampler

Leading principal minor analysis

Consider the following relationship between the leading principal minors of order p and $(p-1)$ between iteration k and $k+1$ for Bayesian graphical lasso block Gibbs sampler in 2.3. Consider the case when $p = 2$

$$\begin{aligned}
\gamma &= \omega_{22} - \omega_{21}\omega_{11}^{-1}\omega_{12} \\
\det(\boldsymbol{\Omega_{11}}^{(k)}) &= \omega_{11} \\
\det(\boldsymbol{\Omega}^{(k+1)}) &= \det\begin{pmatrix} \omega_{11}^{(k)} & \omega_{12}^{(k+1)} \\ \omega_{21}^{(k+1)} & \omega_{22}^{(k+1)} \end{pmatrix} \\
&= \omega_{11}\omega_{22} - \omega_{12}\omega_{21} \\
&= \gamma \det(\boldsymbol{\Omega_{11}}^{(k)}).
\end{aligned}$$

© University of Pretoria

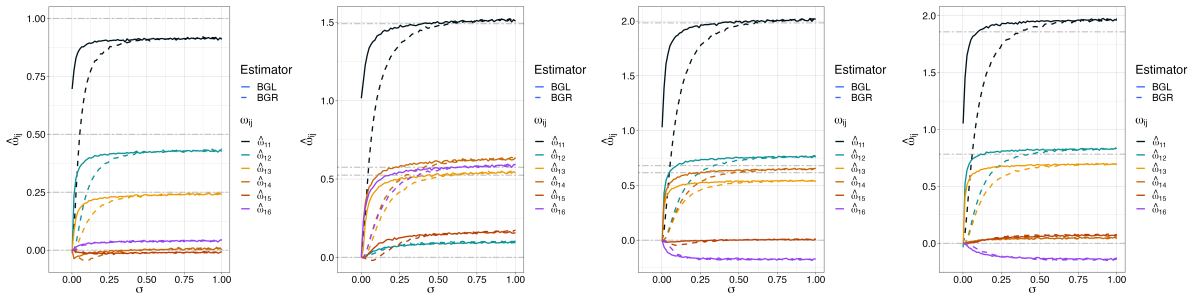
© University of Pretoria

Appendix B

Bayesian Graphical Modeling with a Ridge Prior

B.1 Block Gibbs sampler

This section contains additional figures of the solution paths in Section 3.2 of of Chapter 3.

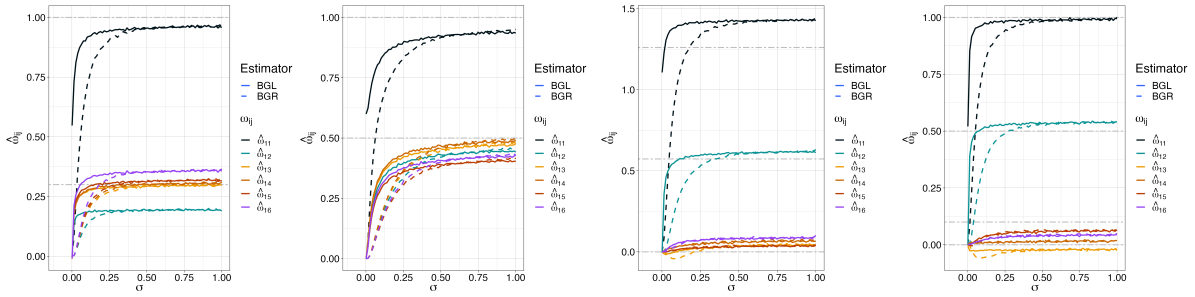


(a) AR(2) (b) random (c) scale-free (d) hub

Fig. B.1 *Coefficient paths for first row of elements in varying* $\mathbf{\Omega}$ *topologies.* Each panel displays the convergence path for the Bayesian graphical lasso (dashed), Bayesian graphical ridge type I (dashed) and Bayesian graphical ridge type II (dotted) for $n > p$ and varying σ. The horizontal grey lines (dot-dash) indicate the true values of ω_{ij}.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

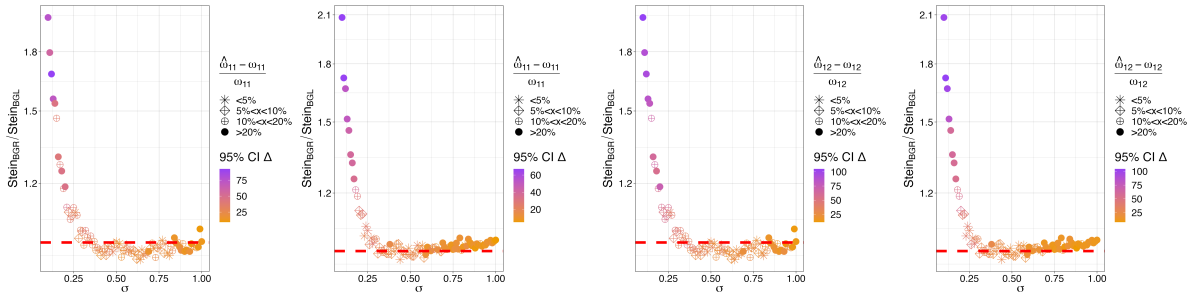


(a) star (b) full (c) band (d) circle

Fig. B.2 *Coefficient paths for first row of elements in varying* $\mathbf{\Omega}$ *topologies.* Each panel displays the convergence path for the Bayesian graphical lasso (dashed), Bayesian graphical ridge type I (dashed) and Bayesian graphical ridge type II (dotted) for $n > p$ and varying σ. The horizontal grey lines (dot-dash) indicate the true values of ω_{ij}.

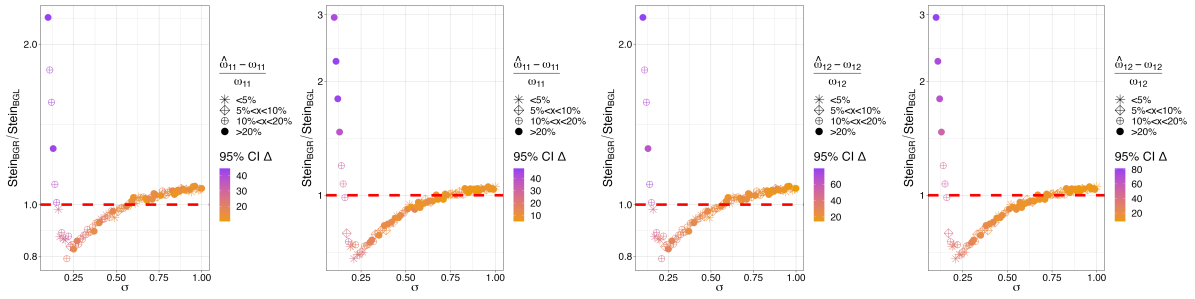
B.2 Bayesian precision matrix inference

This section contains additional figures of the Stein's loss comparison in Section 3.3 of Chapter 3.



(a) ω_{11}, $p/n = 0.5$ (b) ω_{11}, $p/n = 1$ (c) ω_{12}, $p/n = 0.5$ (d) ω_{12}, $p/n = 1$

Fig. B.3 *Ratio of Stein's loss across a varying penalty parameter for the AR(1) network topology.*

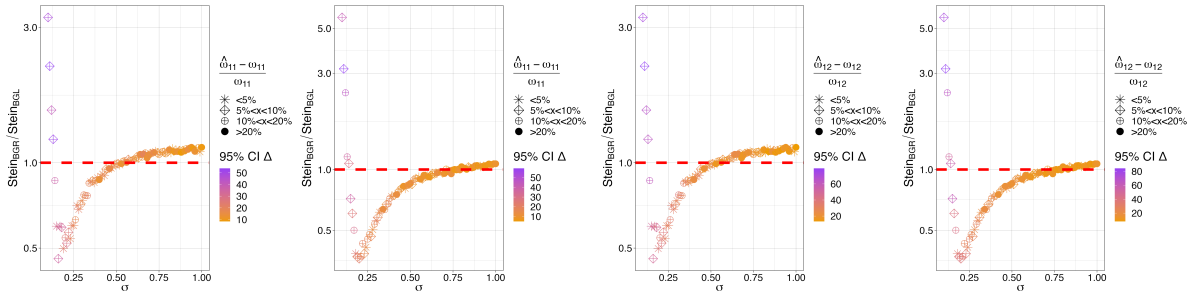


(a) ω_{11}, $p/n = 0.5$ (b) ω_{11}, $p/n = 1$ (c) ω_{12}, $p/n = 0.5$ (d) ω_{12}, $p/n = 1$

Fig. B.4 *Ratio of Stein's loss across a varying penalty parameter for the cluster network topology.*

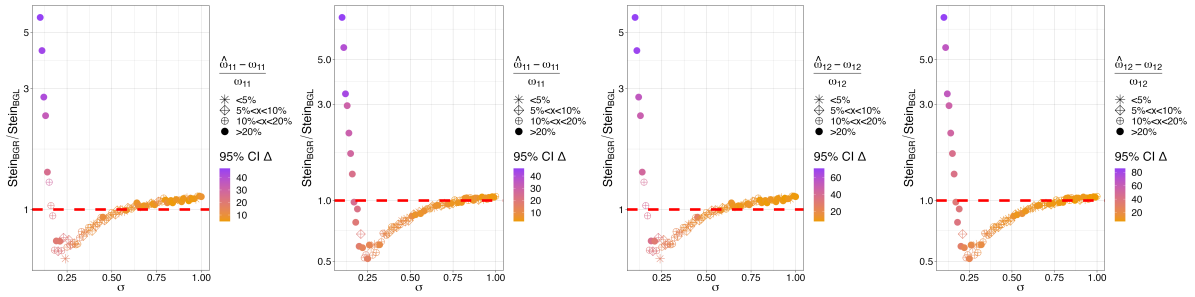
© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



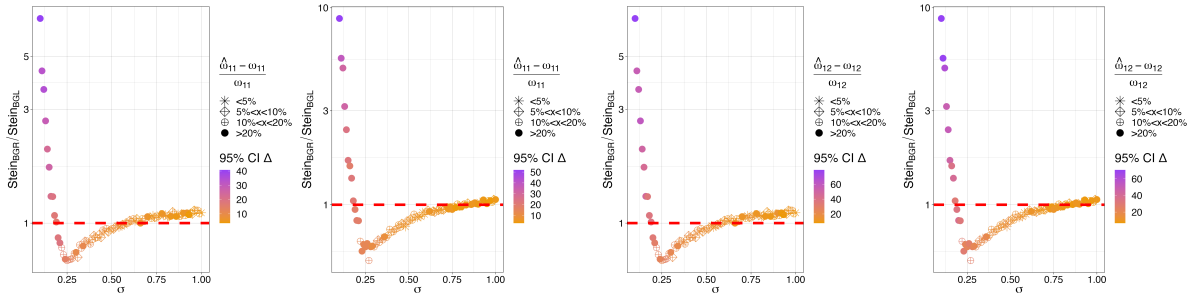
(a) ω_{11}, $p/n = 0.5$ (b) ω_{11}, $p/n = 1$ (c) ω_{12}, $p/n = 0.5$ (d) ω_{12}, $p/n = 1$

Fig. B.5 *Ratio of Stein's loss across a varying penalty parameter for the star network topology.*



(a) ω_{11}, $p/n = 0.5$ (b) ω_{11}, $p/n = 1$ (c) ω_{12}, $p/n = 0.5$ (d) ω_{12}, $p/n = 1$

Fig. B.6 *Ratio of Stein's loss across a varying penalty parameter for the scale-free network topology.*

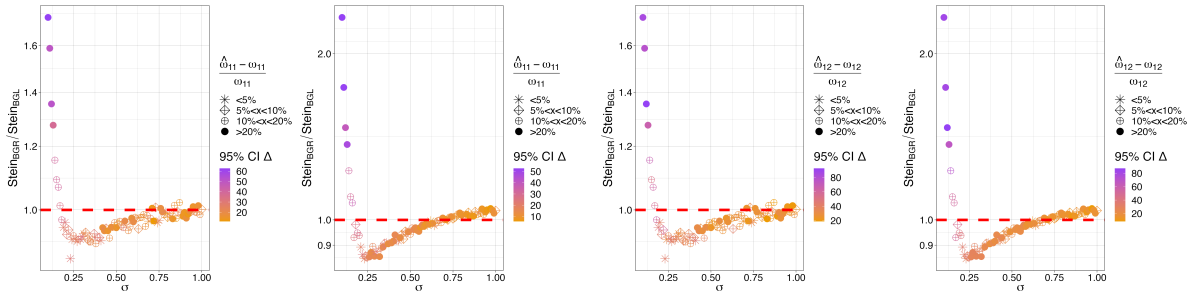


(a) ω_{11}, $p/n = 0.5$ (b) ω_{11}, $p/n = 0.9375$ (c) ω_{12}, $p/n = 0.5$ (d) ω_{12}, $p/n = 1$

Fig. B.7 *Ratio of Stein's loss across a varying penalty parameter for the hub network topology.*

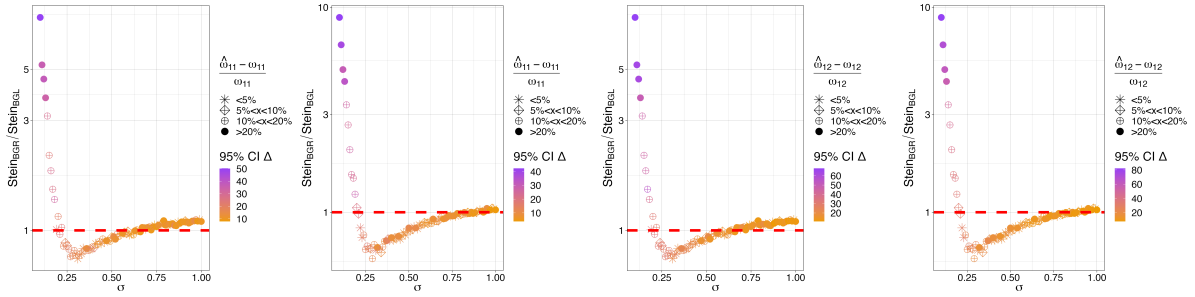
© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



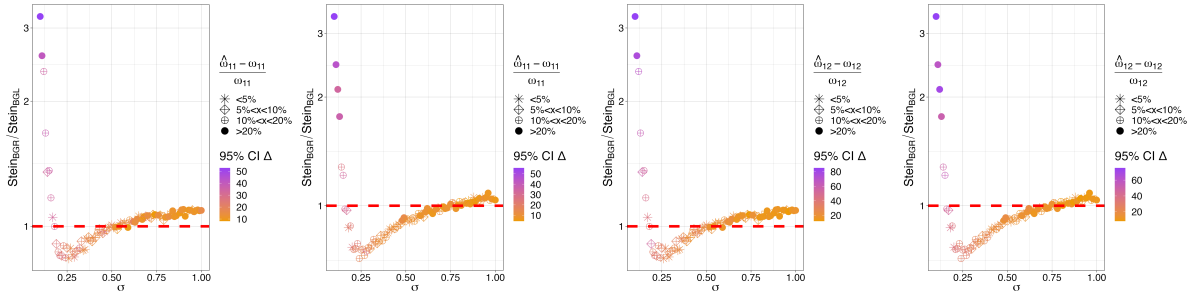
(a) ω_{11}, $p/n = 0.5$ (b) ω_{11}, $p/n = 1$ (c) ω_{12}, $p/n = 0.5$ (d) ω_{12}, $p/n = 1$

Fig. B.8 *Ratio of Stein's loss across a varying penalty parameter for the full network topology.*



(a) ω_{11}, $p/n = 0.5$ (b) ω_{11}, $p/n = 1$ (c) ω_{12}, $p/n = 0.5$ (d) ω_{12}, $p/n = 1$

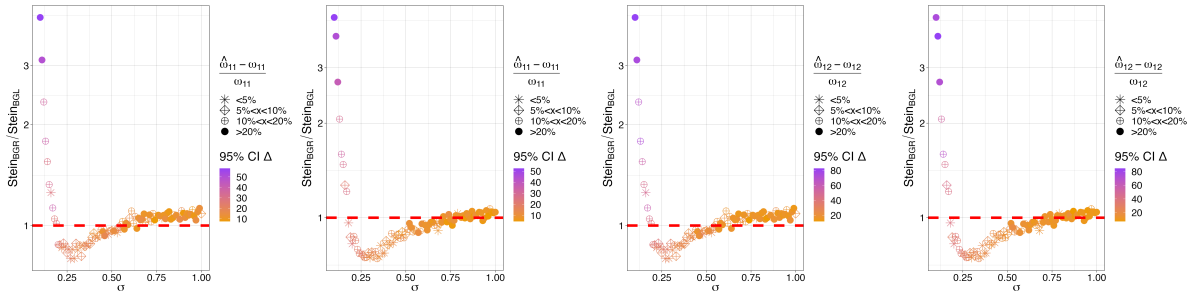
Fig. B.9 *Ratio of Stein's loss across a varying penalty parameter for the band network topology.*



(a) ω_{11}, $p/n = 0.5$ (b) ω_{11}, $p/n = 1$ (c) ω_{12}, $p/n = 0.5$ (d) ω_{12}, $p/n = 1$

Fig. B.10 *Ratio of Stein's loss across a varying penalty parameter for the AR(2) network topology.*

© University of Pretoria



(a) ω_{11}, $p/n = 0.5$ (b) ω_{11}, $p/n = 1$ (c) ω_{12}, $p/n = 0.5$ (d) ω_{12}, $p/n = 1$

Fig. B.11 *Ratio of Stein's loss across a varying penalty parameter for the circle network topology.*

B.3 Bayesian structure learning

This section contains the remaining tables from Section 3.6 of Chapter 3, which summarise the sensitivity, specificity, F1, and MCC scores across multiple network topologies, Bayesian estimators and edge selection techniques TAI, TAII, 50CI, 90CI and 99CI respectively for $p/n = 1$ and $p/n = 10$.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table B.1 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 1$. The median loss values reported here are based on 30 replications for the BAGR estimator. The cell colour of the best performer is set to grey.

$\frac{p}{n} = 1$	**AR(1)**					**AR(2)**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.92	0.97	1	0.92	0.91	0.23	0.75	0.96	0.28	0.2
Specificity	0.1	0.1	0.1	0.1	0.1	0.98	0.96	0.94	0.94	0.93
F1	0.56	0.18	0.12	0.28	0.4	0.37	0.66	0.57	0.35	0.28
MCC	0.03	0.08	0.08	0.03	0.02	0.34	0.63	0.6	0.27	0.19
$\frac{p}{n} = 1$	**star**					**circle**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.15	0.7	0.98	0.22	0.16	0.19	0.76	0.97	0.38	0.26
Specificity	0.97	0.96	0.96	0.96	0.97	1	1	1	1	1
F1	0.24	0.49	0.5	0.29	0.25	0.31	0.86	0.97	0.54	0.41
MCC	0.22	0.49	0.57	0.25	0.21	0.37	0.86	0.96	0.56	0.45
$\frac{p}{n} = 1$	**scale-free**					**random**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.19	0.77	1	0.24	0.18	0.25	0.87	1	0.29	0.21
Specificity	0.99	0.97	0.96	0.97	0.97	0.98	0.95	0.94	0.95	0.95
F1	0.3	0.54	0.51	0.32	0.27	0.39	0.55	0.43	0.35	0.3
MCC	0.31	0.55	0.57	0.28	0.24	0.38	0.57	0.51	0.29	0.24
$\frac{p}{n} = 1$	**cluster**					**hub**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.93	0.98	1	0.94	0.93	0.2	0.79	1	0.25	0.18
Specificity	0.08	0.08	0.08	0.08	0.08	0.99	0.97	0.96	0.97	0.97
F1	0.49	0.08	0.04	0.23	0.33	0.33	0.58	0.52	0.34	0.28
MCC	0.03	0.04	0.04	0.02	0.02	0.36	0.58	0.57	0.31	0.26
$\frac{p}{n} = 1$	**band**					**full**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.25	0.93	1	0.29	0.2	1	1	1	1	1
Specificity	1	0.98	0.97	0.97	0.97	0	0	0	0	0
F1	0.4	0.82	0.66	0.39	0.3	0.57	0.12	0.05	0.24	0.35
MCC	0.44	0.81	0.69	0.37	0.29	0.02	0.01	0	0.01	0.01

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table B.2 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 1$. The median loss values reported here are based on 30 replications for the BAGL estimator. The cell colour of the best performer is set to grey.

$\frac{p}{n} = 1$	**AR(1)**					**AR(2)**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.96	0.99	1	0.96	0.95	0.6	0.87	0.93	0.56	0.46
Specificity	0.1	0.1	0.1	0.1	0.1	0.96	0.95	0.94	0.95	0.95
F1	0.2	0.13	0.11	0.17	0.22	0.63	0.64	0.61	0.56	0.5
MCC	0.07	0.08	0.08	0.06	0.05	0.58	0.65	0.62	0.51	0.45
$\frac{p}{n} = 1$	**star**					**circle**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.39	0.78	0.92	0.34	0.26	0.06	0.06	0.06	0.11	0.06
Specificity	0.96	0.96	0.96	0.96	0.96	1	1	0.99	0.99	1
F1	0.4	0.48	0.49	0.36	0.32	0.11	0.12	0.12	0.2	0.11
MCC	0.36	0.5	0.54	0.32	0.28	0.01	0.03	0.04	0.23	0.01
$\frac{p}{n} = 1$	**scale-free**					**random**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.46	0.78	0.9	0.42	0.32	0.54	0.84	0.92	0.46	0.37
Specificity	0.97	0.97	0.96	0.97	0.97	0.96	0.95	0.95	0.96	0.96
F1	0.51	0.58	0.56	0.46	0.41	0.54	0.56	0.52	0.48	0.45
MCC	0.48	0.58	0.59	0.42	0.38	0.5	0.57	0.56	0.44	0.4
$\frac{p}{n} = 1$	**cluster**					**hub**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.94	0.99	1	0.95	0.94	0.48	0.78	0.9	0.41	0.33
Specificity	0.08	0.08	0.08	0.08	0.08	0.98	0.97	0.97	0.97	0.98
F1	0.12	0.06	0.0.06	0.12	0.17	0.56	0.62	0.61	0.49	0.44
MCC	0.02	0.04	0.04	0.03	0.02	0.53	0.62	0.63	0.46	0.43
$\frac{p}{n} = 1$	**band**					**full**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.59	0.89	0.96	0.54	0.42	1	1	1	1	1
Specificity	0.99	0.98	0.98	0.99	0.99	0	0	0	0	0
F1	0.69	0.8	0.77	0.64	0.55	0.12	0.05	0.04	0.12	0.16
MCC	0.68	0.79	0.78	0.62	0.55	0.01	0	0	0.01	0.01

Table B.3 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 1$. The median loss values reported here are based on 30 replications for the BAGL estimator. The cell colour of the best performer is set to grey.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table B.4 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 10$. The median loss values reported here are based on 30 replications for the BAGR estimator. The cell colour of the best performer is set to grey.

$\frac{p}{n} = 10$	**AR(1)**					**AR(2)**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.94	1	1	0.9	0.9	0.48	1	1	0.12	0.11
Specificity	0.1	0.1	0.1	0.1	0.1	0.92	0.92	0.92	0.93	0.92
F1	0.16	0.04	0.04	0.73	0.81	0.34	0.34	0.34	0.2	0.19
MCC	0.04	0.05	0.05	0.01	0	0.3	0.43	0.43	0.08	0.05
$\frac{p}{n} = 10$	**star**					**circle**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.74	1	1	0.09	0.07	0.32	1	1	0.09	0.07
Specificity	0.96	0.96	0.96	0.96	0.96	0.98	0.96	0.96	0.99	0.99
F1	0.48	0.5	0.5	0.15	0.13	0.45	0.51	0.49	0.16	0.14
MCC	0.5	0.57	0.57	0.1	0.07	0.43	0.57	0.56	0.16	0.12
$\frac{p}{n} = 10$	**scale-free**					**random**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.73	1	1	0.09	0.08	0.76	1	1	0.11	0.1
Specificity	0.96	0.96	0.96	0.96	0.96	0.94	0.94	0.94	0.94	0.94
F1	0.47	0.5	0.5	0.15	0.14	0.4	0.4	0.4	0.18	0.17
MCC	0.49	0.57	0.57	0.1	0.08	0.42	0.48	0.48	0.09	0.07
$\frac{p}{n} = 10$	**cluster**					**hub**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.98	1	1	0.93	0.92	0.75	1	1	0.09	0.08
Specificity	0.08	0.08	0.08	0.08	0.08	0.96	0.96	0.96	0.97	0.97
F1	0.08	0.04	0.04	0.64	0.73	0.48	0.51	0.51	0.16	0.14
MCC	0.05	0.04	0.04	0.02	0.01	0.5	0.57	0.57	0.12	0.1
$\frac{p}{n} = 10$	**band**					**full**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.79	1	1	0.09	0.08	1	1	1	1	1
Specificity	0.96	0.96	0.96	0.96	0.96	0	0	0	0	0
F1	0.5	0.5	0.5	0.16	0.14	0.11	0.04	0.04	0.73	0.81
MCC	0.52	0.57	0.57	0.11	0.08	0	0	0	0.02	0.03

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table B.5 *Summary of Sensitivity, Specificity, F1 and MCC performance across multiple network topologies and edge selection methods for* $p/n = 10$. The median loss values reported here are based on 30 replications for the BAGL estimator. The cell colour of the best performer is set to grey.

$\frac{p}{n} = 10$	**AR(1)**					**AR(2)**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.94	0.97	1	0.91	0.91	0.42	0.81	0.92	0.14	0.12
Specificity	0.1	0.1	0.1	0.1	0.1	0.92	0.92	0.92	0.92	0.92
F1	0.14	0.06	0.05	0.5	0.57	0.31	0.34	0.33	0.21	0.2
MCC	0.03	0.04	0.05	0.01	0	0.27	0.39	0.41	0.1	0.08
$\frac{p}{n} = 10$	**star**					**circle**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.47	0.92	1	0.09	0.08	0.41	0.72	0.8	0.15	0.12
Specificity	0.96	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.98	0.98
F1	0.41	0.49	0.5	0.15	0.14	0.49	0.58	0.57	0.25	0.21
MCC	0.38	0.54	0.57	0.09	0.08	0.46	0.57	0.57	0.25	0.21
$\frac{p}{n} = 10$	**scale-free**					**random**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.49	0.92	1	0.1	0.09	0.5	0.92	1	0.12	0.11
Specificity	0.96	0.96	0.96	0.96	0.96	0.94	0.94	0.94	0.94	0.94
F1	0.42	0.49	0.5	0.17	0.15	0.36	0.4	0.4	0.2	0.18
MCC	0.39	0.54	0.57	0.12	0.1	0.33	0.47	0.48	0.12	0.1
$\frac{p}{n} = 10$	**cluster**					**hub**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.97	1	1	0.92	0.92	0.49	0.92	1	0.1	0.09
Specificity	0.08	0.08	0.08	0.08	0.07	0.96	0.96	0.96	0.96	0.96
F1	0.1	0.05	0.04	0.5	0.57	0.42	0.5	0.51	0.17	0.15
MCC	0.04	0.04	0.04	0	0	0.39	0.55	0.57	0.13	0.11
$\frac{p}{n} = 10$	**band**					**full**				
	50CI	90CI	99CI	TAI	TAII	50CI	90CI	99CI	TAI	TAII
Sensitivity	0.53	0.96	1	0.11	0.09	1	1	1	1	1
Specificity	0.96	0.96	0.96	0.96	0.96	0	0	0	0	0
F1	0.44	0.5	0.5	0.18	0.16	0.12	0.05	0.04	0.52	0.59
MCC	0.42	0.56	0.57	0.15	0.13	0.01	0	0	0.01	0.02

© University of Pretoria

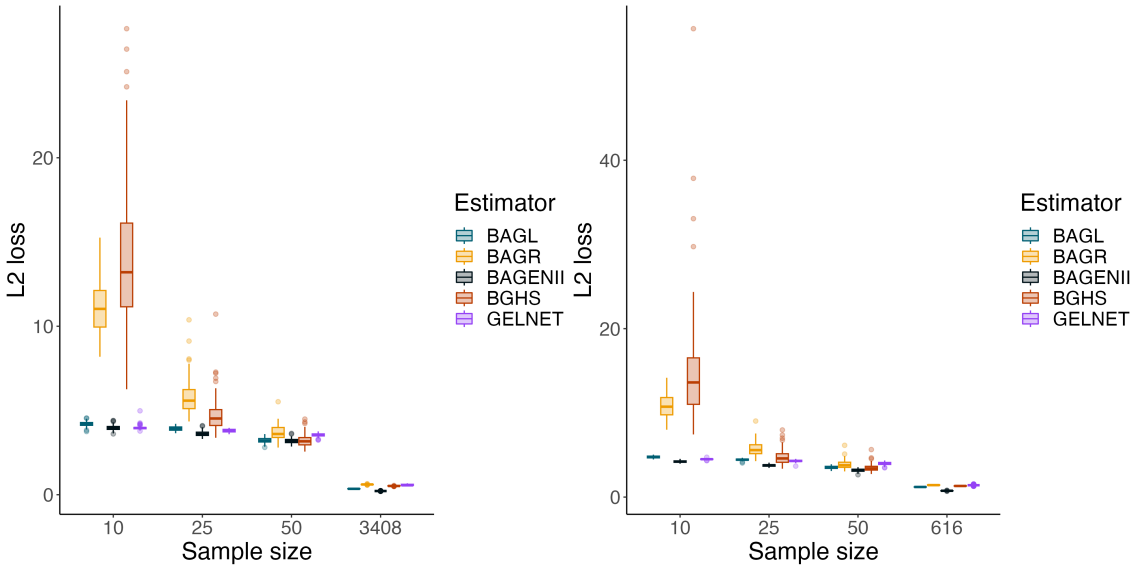
© University of Pretoria

Appendix C

Bayesian Graphical Modeling with a Naïve Elastic Net Prior

C.1 SEER breast cancer

This section contains additional metadata of the SEER breast cancer data, and figures of the comparison between the BAGL, BAR and BAGEN estimators in Section 4.5 of Chapter 4.



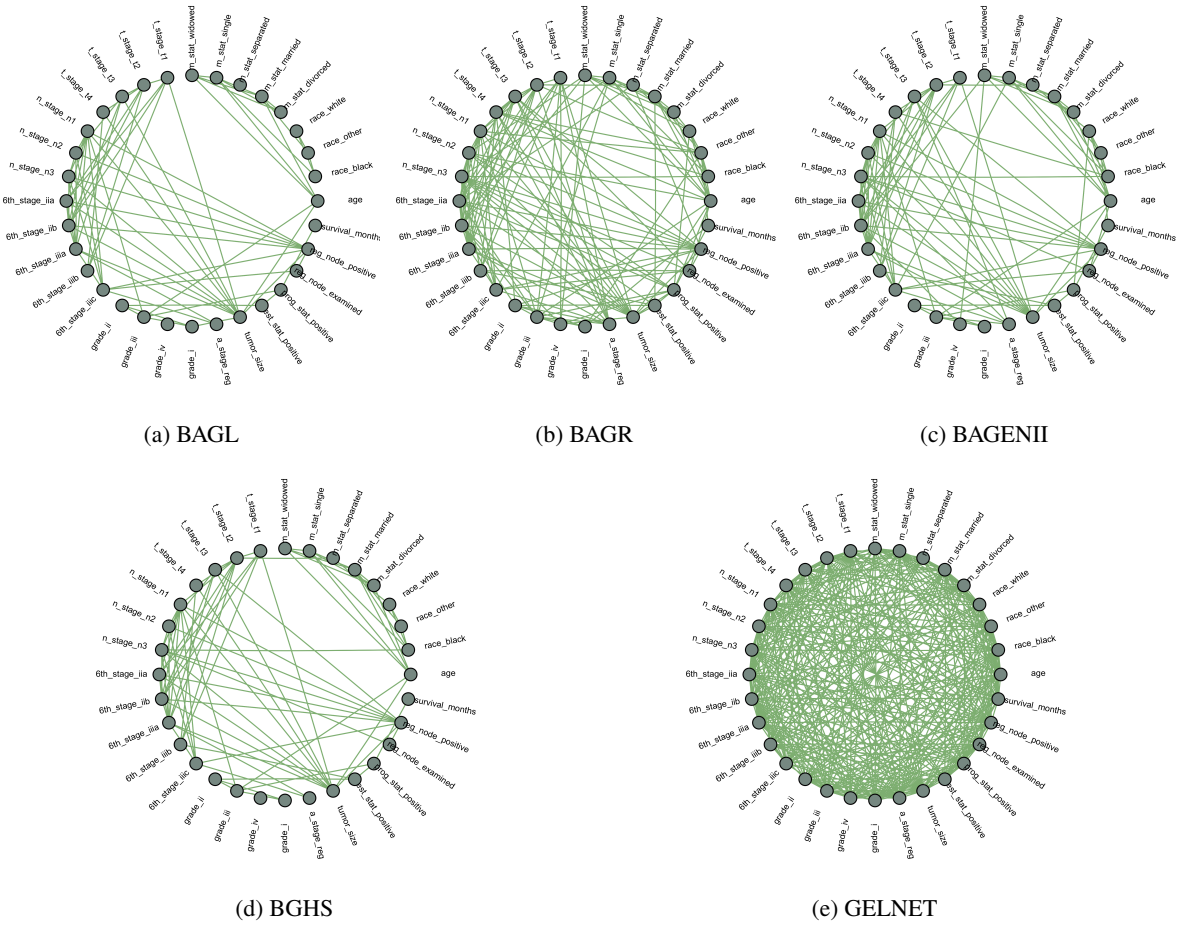
(a) alive

(b) deceased

Fig. C.1 *Loss comparison between the BAGL, BAGR, BAGENII, BGHS and GELNET on the SEER breast cancer data.* The left-hand panel illustrates the Frobenius loss (referred to as L2 loss on the y-axis) on the cohort of patients who survived the disease and the right-hand panel depicts Frobenius loss for the the cohort of patients who succumbed to the disease.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



(a) BAGL (b) BAGR (c) BAGENII

(d) BGHS (e) GELNET

Fig. C.2 *The inferred networks from the BAGL, BAGR, BAGENII, BGHS and GELNET for the SEER breast cancer data for alive patients.* The Bayesian networks are determined using a 90% symmetric credible interval.

C.1.1 SEER breast cancer metadata

The SEER breast cancer metadata document provided by (Teng, 2019) is included here for smoothness of reading. This dissertation did not contribute to the collection nor summarisation of the data.

© University of Pretoria

SEER Breast Cancer Dataset

Abstract: This dataset of breast cancer patients was obtained from the 2017 November update of the SEER Program of the NCI, which provides information on population-based cancer statistics. The dataset involved female patients with infiltrating duct and lobular carcinoma breast cancer (SEER primary cites recode NOS histology codes 8522/3) diagnosed in 2006-2010. Patients with unknown tumor size, examined regional LNs, regional positive LNs, and patients whose survival months were less than 1 month were excluded; thus, 4024 patients were ultimately included.

AGE

SEER*Stat Name: Age at diagnosis

Field Description: This data item represents the age of the patient at diagnosis **for this cancer**. The code represents the patient's actual age in years.

RACE

SEER*Stat Name: Race recode (White, Black, Other)

Field Description: Race recode is based on the race variables and the American Indian/Native American IHS link variable. This recode should be used to link to the populations for white, black and other. It is independent of Hispanic ethnicity. For more information, see : http://seer.cancer.gov/seerstat/variables/seer/race_ethnicity.

1	White
2	Black
3	Other (American Indian/AK Native, Asian/Pacific Islander)
4	Other unspecified (1991+)
5	Unknown

MARITAL STATUS

SEER*Stat Name: Marital status at diagnosis

Field Description: This data item identifies the patient's marital status at the time of diagnosis for the reportable tumor.

1	Single (never married)
2	Married (including common law)
3	Separated
4	Divorced
5	Widowed

T STAGE

SEER*Stat Name: Breast Adjusted AJCC 6th T (1988+)

Field Description: Created from merged EOD 3rd Edition and Collaborative Stage disease

© University of Pretoria

information. Currently only available for Breast schema. For more information see http://seer.cancer.gov/seerstat/variables/seer/ajcc-stage/6th.

1	T1
2	T2
3	T3
4	T4

N STAGE

SEER*Stat Name: Breast Adjusted AJCC 6th N (1988+)

Field Description: Created from merged EOD 3rd Edition and Collaborative Stage disease information. Currently only available for Breast schema. For more information see http://seer.cancer.gov/seerstat/variables/seer/ajcc-stage/6th.

1	N1
2	N2
3	N3

6TH STAGE

SEER*Stat Name: Breast Adjusted AJCC 6th Stage (1988+)

Field Description: Created from merged EOD 3rd Edition and Collaborative Stage disease information. Currently only available for Breast schema. For more information see http://seer.cancer.gov/seerstat/variables/seer/ajcc-stage/6th.

1	IIA
2	IIB
3	IIIA
4	IIIB
5	IIIC

GRADE

SEER*Stat Name: Grade

Field Description: Grading and differentiation codes of 1-4, 9 are defined in ICD-O-2; 1992. Grade information may be incomplete for cases diagnosed before 1977. In the early 1980's, additional codes specifying T-cell, B-cell, or null cell involvement in lymphomas and leukemias (histologies M9590-9940) were introduced by SEER. Because the reporting requirements and medical terminology have changed over time, care should be exercised when analyzing this information.

1	Grade I; grade i; grade 1; well differentiated; differentiated, NOS
2	Grade II; grade ii; grade 2; moderately differentiated; moderately differentiated; intermediate differentiation
3	Grade III; grade iii; grade 3; poorly differentiated; differentiated
4	Grade IV; grade iv; grade 4; undifferentiated; anaplastic

A STAGE

SEER*Stat Name: SEER historic stage A

© University of Pretoria

Field Description: Derived from Collaborative Stage (CS) for 2004+ and Extent of Disease (EOD) from 1973-2003. It is a simplified version of stage: in situ, localized, regional, distant, & unknown. Over time several different EOD schemes have been used. Thus caution should be used when doing trend analysis. For more information including sites and years for which it isn't calculated, see http://seer.cancer.gov/seerstat/variables/seer/lrd-stage.

1	Regional — A neoplasm that has extended 1) beyond the limits of the organ of origin directly into surrounding organs or tissues; 2) into regional lymph nodes by way of the lymphatic system; or 3) by a combination of extension and regional lymph nodes.
2	Distant — A neoplasm that has spread to parts of the body remote from the primary tumor either by direct extension or by discontinuous metastasis (e.g., implantation or seeding) to distant organs, issues, or via the lymphatic system to distant lymph nodes.

TUMOR SIZE

SEER*Stat Name: CS tumor size (2004+)

Field Description: Information on tumor size. Available for 2004+. Earlier cases may be converted and new codes added which weren't available for use prior to the current version of CS. Each indicates exact size in millimeters. For more information, see http://seer.cancer.gov/seerstat/variables/seer/ajcc-stage.

ESTROGEN STATUS

SEER*Stat Name: ER Status Recode Breast Cancer (1990+)

Field Description: Created by combining information from Tumor marker 1 (1990-2003) (NAACCR Item #=1150), with information from CS site-specific factor 1 (2004+) (NAACCR Item #=2880). This field is blank for non-breast cases and cases diagnosed before 1990.

1	Positive
2	Negative

PROGESTERONE STATUS

SEER*Stat Name: PR Status Recode Breast Cancer (1990+)

Field Description: Created by combining information from Tumor marker 2 (1990-2003) (NAACCR Item #=1150), with information from CS site-specific factor 2 (2004+) (NAACCR Item #=2880). This field is blank for non-breast cases and cases diagnosed before 1990.

1	Positive
2	Negative

REGIONAL NODES EXAMINED

SEER*Stat Name: Regional nodes examined (1988+)

Field Description: Records the total number of regional lymph nodes that were removed and examined by the pathologist.

REGIONAL NODES POSITIVE

SEER*Stat Name: Regional nodes positive (1988+)

© University of Pretoria

Field Description: Records the exact number of regional lymph nodes examined by the pathologist that were found to contain metastases.

SURVIVAL MONTHS

SEER*Stat Name: Survival Months

Field Description: Created using complete dates, including days, therefore may differ from survival time calculated from year and month only. For more information, see http://seer.cancer.gov/survivaltime.

STATUS

SEER*Stat Name: Vital status recode (study cutoff used)

Field Description: Any patient that dies after the follow-up cut-off date is recoded to alive as of the cut-off date.

1	Alive
2	Dead

© University of Pretoria

Appendix D

Differential Network Analysis: A Bayesian Perspective

D.1 Structure learning

This section contains additional figures for the structure learning comparison in Subsection 5.2.1 of Chapter 5.

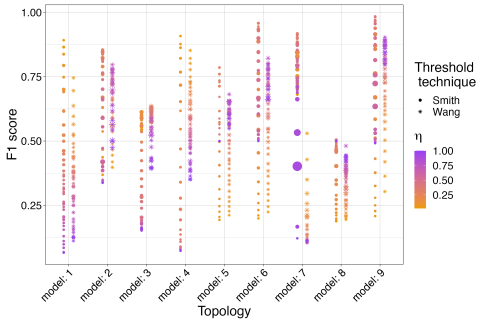


Fig. D.1 *F1 score for various topologies, thresholding techniques and varying thresholds.* The curves illustrate how varying thresholds and Bayesian sparsity criteria in (5.12) and (5.13) affect the F1 score of network recovery for graphs with a dimension of $p = 30$, with point sizes reflecting standard error magnitude.

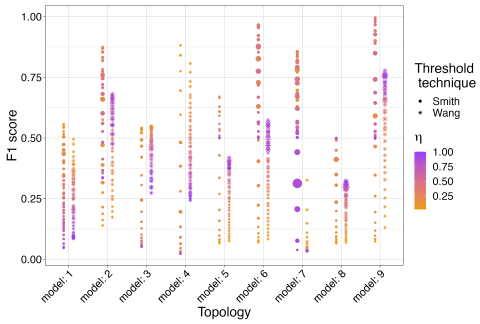
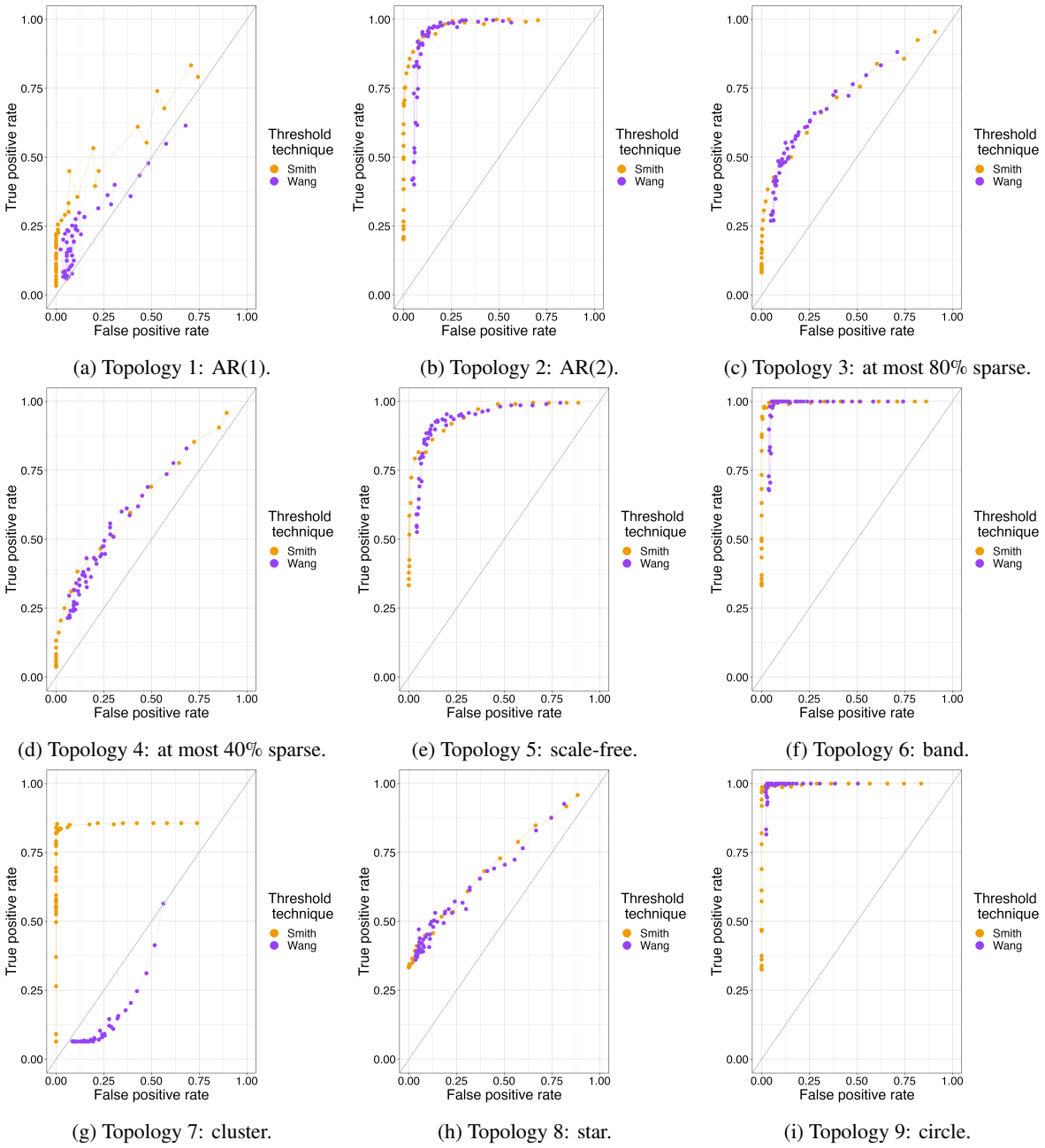


Fig. D.2 *F1 score for various topologies, thresholding techniques and varying thresholds.* The curves illustrate how varying thresholds and Bayesian sparsity criteria in (5.12) and (5.13) affect the F1 score of network recovery for graphs with a dimension of $p = 100$, with point sizes reflecting standard error magnitude.

© University of Pretoria



(a) Topology 1: AR(1). (b) Topology 2: AR(2). (c) Topology 3: at most 80% sparse.

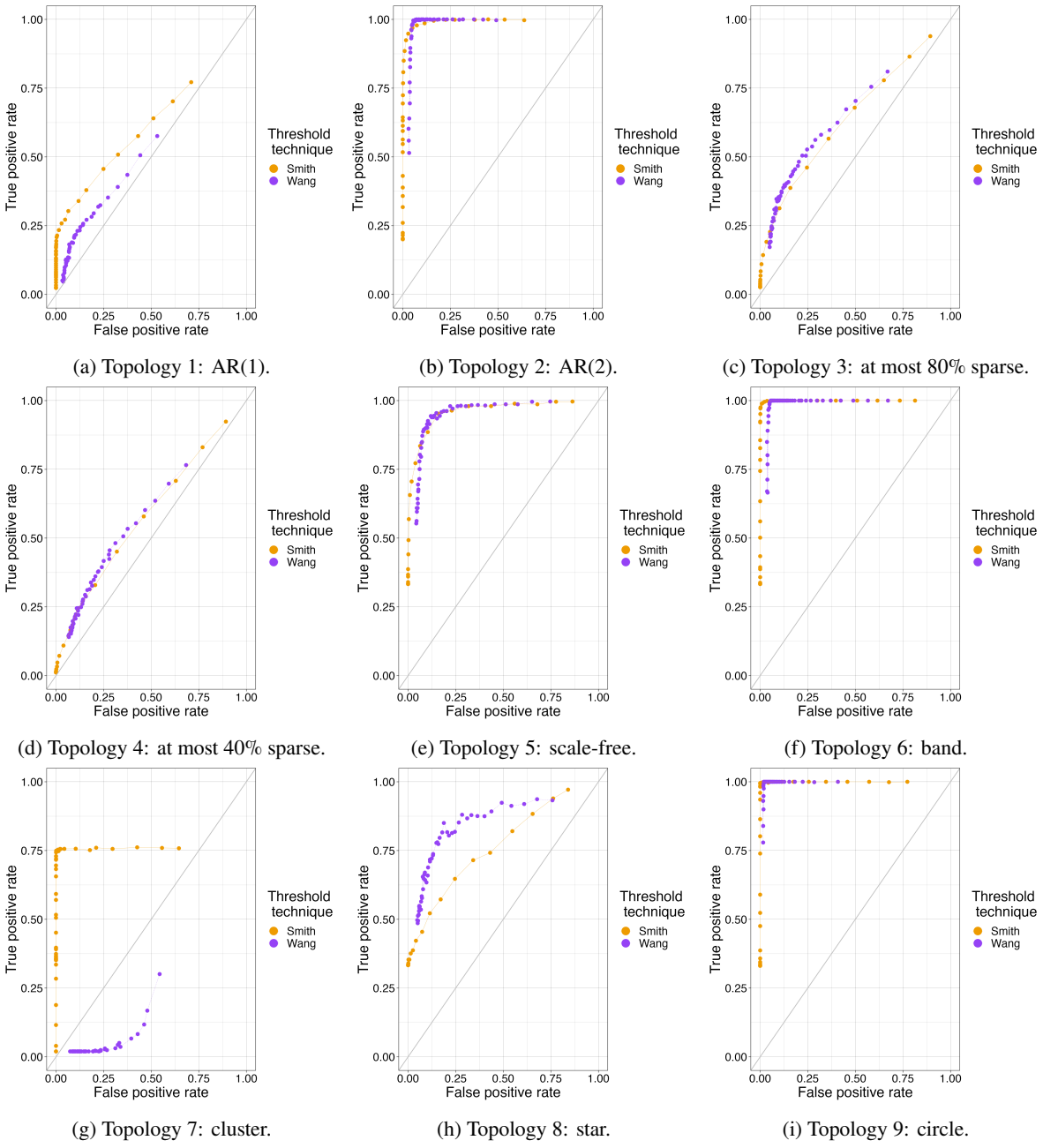
(d) Topology 4: at most 40% sparse. (e) Topology 5: scale-free. (f) Topology 6: band.

(g) Topology 7: cluster. (h) Topology 8: star. (i) Topology 9: circle.

Fig. D.3 *Receiver operating characteristic curves for various topologies thresholding techniques.* The ROC curves display the network recovery performance for various graph structures under varying thresholds for each Bayesian sparsity criterion in (5.12) and (5.13) for dimension $p = 30$.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



(a) Topology 1: AR(1).

(b) Topology 2: AR(2).

(c) Topology 3: at most 80% sparse.

(d) Topology 4: at most 40% sparse.

(e) Topology 5: scale-free.

(f) Topology 6: band.

(g) Topology 7: cluster.

(h) Topology 8: star.

(i) Topology 9: circle.

Fig. D.4 *Receiver operating characteristic curves for various topologies thresholding techniques.* The ROC curves display the network recovery performance for various graph structures under varying thresholds for each Bayesian sparsity criterion in (5.12) and (5.13) for dimension $p = 100$.

© University of Pretoria

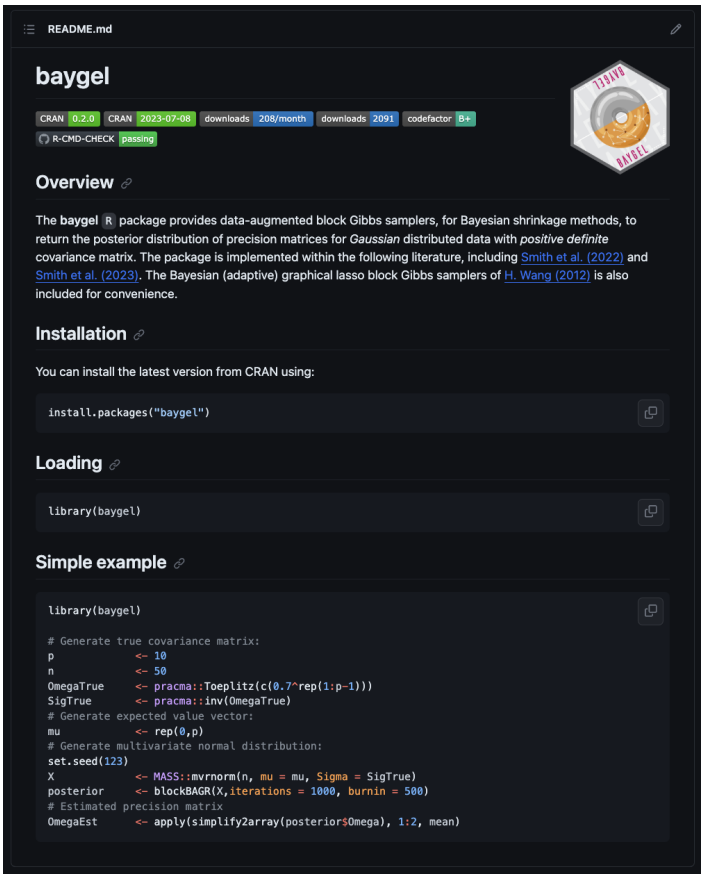
© University of Pretoria

Appendix E

The baygel Package

A key goal of this dissertation is to provide publicly available software. To this end the 'baygel' R package is included. The package offers block Gibbs samplers for the Bayesian graphical lasso, ridge and naïve elastic net techniques discussed in Chapters 2, 3 and 4, respectively. The MCMC samplers are implemented using C++. To facilitate usage and exploration for practitioners the adaptive and non adaptive samplers are included. The R package is available on the Comprehensive R Archive Network and GitHub via baygel.

Finally, the R code with which simulations were conducted in this dissertation is available on GitHub R Simulations: Bayesian Learning of Regularized Gaussian Graphical Networks.



README.md

baygel

CRAN 0.2.0 | CRAN 2023-07-08 | downloads 208/month | downloads 2091 | codefactor B+
R-CMD-CHECK passing

Overview

The **baygel** R package provides data-augmented block Gibbs samplers, for Bayesian shrinkage methods, to return the posterior distribution of precision matrices for *Gaussian* distributed data with *positive definite* covariance matrix. The package is implemented within the following literature, including Smith et al. (2022) and Smith et al. (2023). The Bayesian (adaptive) graphical lasso block Gibbs samplers of H. Wang (2012) is also included for convenience.

Installation

You can install the latest version from CRAN using:

```
install.packages("baygel")
```

Loading

```
library(baygel)
```

Simple example

```
library(baygel)

# Generate true covariance matrix:
p             <- 10
n             <- 50
OmegaTrue     <- pracma::Toeplitz(c(0.7^rep(1:p-1)))
SigTrue       <- pracma::inv(OmegaTrue)
# Generate expected value vector:
mu            <- rep(0,p)
# Generate multivariate normal distribution:
set.seed(123)
X             <- MASS::mvrnorm(n, mu = mu, Sigma = SigTrue)
posterior     <- blockBAGR(X,iterations = 1000, burnin = 500)
# Estimated precision matrix
OmegaEst      <- apply(simplify2array(posterior$Omega), 1:2, mean)
```

© University of Pretoria

Package 'baygel'

November 3, 2023

Type Package

Title Bayesian Shrinkage Estimators for Gaussian Graphical Models

Version 0.3.0

Date 2023-11-02

Description This R package offers block Gibbs samplers for the Bayesian (adaptive) graphical lasso, ridge, and naïve elastic net priors. These samplers facilitate the simulation of the posterior distribution of precision matrices for Gaussian distributed data and were originally proposed by: Wang (2012) <doi:10.1214/12-BA729>; Smith et al. (2022) <doi:10.48550/arXiv.2210.16290> and Smith et al. (2023) <doi:10.48550/arXiv.2306.1419

License GPL (>= 3)

Imports Rcpp (>= 1.0.8),
RcppArmadillo (>= 0.11.1.1.0)

LinkingTo Rcpp,
RcppArmadillo,
RcppProgress

Suggests MASS,
pracma

RoxygenNote 7.2.3

URL https://github.com/Jarod-Smithy/baygel

Encoding UTF-8

R topics documented:

blockBAGENI 2
blockBAGENII 3
blockBAGL 4
blockBAGR 5
blockBGEN 6
blockBGL 7
blockBGR 7

Index **9**

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

`blockBAGENI` *Type I naïve Bayesian adaptive graphical elastic net block Gibbs sampler for Gaussian graphical models.*

Description

Implements the Type I naïve Bayesian adaptive graphical elastic net block Gibbs sampler to simulate the posterior distribution of the precision matrix for Gaussian graphical models.

Usage

```
blockBAGENI(
  X,
  burnin,
  iterations,
  verbose = TRUE,
  r = 0.01,
  s = 1e-06,
  a = 0.5,
  b = 0.5
)
```

Arguments

`X`	Numeric data matrix.
`burnin`	An integer specifying the number of burn-in iterations.
`iterations`	An integer specifying the length of the Markov chain after the burn-in iterations.
`verbose`	A logical determining whether the progress of the MCMC sampler should be displayed.
`r`	A double specifying the value of the shape parameter for the gamma prior associated with the Bayesian graphical lasso penalty term.
`s`	A double specifying the value of the scale parameter for the gamma prior associated with the Bayesian graphical lasso penalty term.
`a`	A double specifying the value of the shape parameter for the inverse gamma prior associated with the Bayesian graphical ridge penalty term.
`b`	A double specifying the value of the scale parameter for the inverse gamma prior associated with the Bayesian graphical ridge penalty term.

Value

A list containing precision ‘Omega‘ and covariance ‘Sigma‘ matrices from the Markov chains.

Examples

```
# Generate true covariance matrix:
p            <- 10
n            <- 50
SigTrue      <- pracma::Toeplitz(c(0.7^rep(1:p-1)))
OmegaTrue    <- pracma::inv(SigTrue)
# Generate expected value vector:
```

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA


```
mu            <- rep(0,p)
# Generate multivariate normal distribution:
set.seed(123)
X             <- MASS::mvrnorm(n, mu = mu, Sigma = SigTrue)
posterior     <- blockBAGENI(X, iterations = 1000, burnin = 500)
```

`blockBAGENII` *Type II naïve Bayesian adaptive graphical elastic net block Gibbs sampler for Gaussian graphical models.*

Description

Implements the Type II naïve Bayesian adaptive graphical elastic net block Gibbs sampler to simulate the posterior distribution of the precision matrix for Gaussian graphical models.

Usage

```
blockBAGENII(X, burnin, iterations, verbose = TRUE, s = 1e-06, b = 0.5)
```

Arguments

`X`	Numeric data matrix.
`burnin`	An integer specifying the number of burn-in iterations.
`iterations`	An integer specifying the length of the Markov chain after the burn-in iterations.
`verbose`	A logical determining whether the progress of the MCMC sampler should be displayed.
`s`	A double specifying the value of the rate parameter for the exponential prior associated with the Bayesian graphical lasso penalty term.
`b`	A double specifying the value of the rate parameter for the exponential prior associated with the Bayesian graphical ridge penalty term.

Value

A list containing precision ‘Omega‘ and covariance ‘Sigma‘ matrices from the Markov chains.

Examples

```
# Generate true covariance matrix:
p             <- 10
n             <- 50
SigTrue       <- pracma::Toeplitz(c(0.7^rep(1:p-1)))
OmegaTrue     <- pracma::inv(SigTrue)
# Generate expected value vector:
mu            <- rep(0,p)
# Generate multivariate normal distribution:
set.seed(123)
X             <- MASS::mvrnorm(n, mu = mu, Sigma = SigTrue)
posterior     <- blockBAGENII(X, iterations = 1000, burnin = 500, b = 1e-3, s = 1e-1)
```

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

blockBAGL	*Bayesian adaptive graphical lasso block Gibbs sampler for Gaussian graphical models.*

Description

Implements a Bayesian adaptive graphical lasso block Gibbs sampler to simulate the posterior distribution of the precision matrix for Gaussian graphical models.

Usage

```
blockBAGL(X, burnin, iterations, verbose = TRUE, r = 0.01, s = 1e-06)
```

Arguments

X	Numeric data matrix.
burnin	An integer specifying the number of burn-in iterations.
iterations	An integer specifying the length of the Markov chain after the burn-in iterations.
verbose	A logical determining whether the progress of the MCMC sampler should be displayed.
r	A double specifying the value of the shape parameter for the gamma prior.
s	A double specifying the value of the scale parameter for the gamma prior.

Value

A list containing precision ‘Omega‘ and covariance ‘Sigma‘ matrices from the Markov chains.

Examples

```
# Generate true covariance matrix:
p             <- 10
n             <- 50
SigTrue       <- pracma::Toeplitz(c(0.7^rep(1:p-1)))
OmegaTrue     <- pracma::inv(SigTrue)
# Generate expected value vector:
mu            <- rep(0,p)
# Generate multivariate normal distribution:
set.seed(123)
X             <- MASS::mvrnorm(n, mu = mu, Sigma = SigTrue)
posterior     <- blockBAGL(X, iterations = 1000, burnin = 500)
```

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

`blockBAGR` *Bayesian adaptive graphical ridge block Gibbs sampler for Gaussian graphical models.*

Description

Implements a Bayesian adaptive graphical ridge block Gibbs sampler to simulate the posterior distribution of the precision matrix for Gaussian graphical models.

Usage

```
blockBAGR(X, burnin, iterations, verbose = TRUE, a = 0.5, b = 0.5)
```

Arguments

`X`	Numeric data matrix.
`burnin`	An integer specifying the number of burn-in iterations.
`iterations`	An integer specifying the length of the Markov chain after the burn-in iterations.
`verbose`	A logical determining whether the progress of the MCMC sampler should be displayed.
`a`	A double specifying the value of the shape parameter for the inverse gamma prior.
`b`	A double specifying the value of the scale parameter for the inverse gamma prior.

Value

A list containing precision ‘Omega‘ and covariance ‘Sigma‘ matrices from the Markov chains.

Examples

```
# Generate true covariance matrix:
p             <- 10
n             <- 50
SigTrue       <- pracma::Toeplitz(c(0.7^rep(1:p-1)))
OmegaTrue     <- pracma::inv(SigTrue)
# Generate expected value vector:
mu            <- rep(0,p)
# Generate multivariate normal distribution:
set.seed(123)
X             <- MASS::mvrnorm(n, mu = mu, Sigma = SigTrue)
posterior     <- blockBAGR(X, iterations = 1000, burnin = 500, a = 1, b = 1e-2)
```

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

`blockBGEN` *Naïve Bayesian graphical elastic net block Gibbs sampler for Gaussian graphical models.*

Description

Implements the Bayesian graphical elastic net block Gibbs sampler to simulate the posterior distribution of the precision matrix for Gaussian graphical models.

Usage

```
blockBGEN(X, burnin, iterations, lambda = 1, sig = 1, verbose = TRUE)
```

Arguments

`X`	Numeric data matrix.
`burnin`	An integer specifying the number of burn-in iterations.
`iterations`	An integer specifying the length of the Markov chain after the burn-in iterations.
`lambda`	A numeric value representing the scale (rate) parameter for the double exponential and exponential prior associated with the Bayesian graphical lasso penalty term.
`sig`	A numeric value representing the standard deviation parameter for the double Gaussian and truncated Gaussian prior associated with the Bayesian graphical ridge penalty term.
`verbose`	A logical determining whether the progress of the MCMC sampler should be displayed.

Value

A list containing precision ‘Omega‘ and covariance ‘Sigma‘ matrices from the Markov chains.

Examples

```
# Generate true covariance matrix:
p             <- 10
n             <- 50
SigTrue       <- pracma::Toeplitz(c(0.7^rep(1:p-1)))
OmegaTrue     <- pracma::inv(SigTrue)
# Generate expected value vector:
mu            <- rep(0,p)
# Generate multivariate normal distribution:
set.seed(123)
X             <- MASS::mvrnorm(n, mu = mu, Sigma = SigTrue)
posterior     <- blockBGEN(X, iterations = 1000, burnin = 500, lambda = 1, sig = 1)
```

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

`blockBGL` *Bayesian graphical lasso block Gibbs sampler for Gaussian graphical models.*

Description

Implements a Bayesian graphical lasso block Gibbs sampler to simulate the posterior distribution of the precision matrix for Gaussian graphical models.

Usage

```
blockBGL(X, burnin, iterations, lambda = 1, verbose = TRUE)
```

Arguments

`X`	A numeric matrix.
`burnin`	An integer representing the number of burn-in iterations.
`iterations`	An integer representing the length of the Markov chain post burn-in.
`lambda`	A numeric value representing the scale (rate) parameter for the double exponential and exponential prior.
`verbose`	A logical indicating if the MCMC sampler progress should be printed.

Value

A list containing precision 'Omega' and covariance 'Sigma' matrices from the Markov chains.

Examples

```
# Generate true covariance matrix:
p             <- 10
n             <- 50
SigTrue       <- pracma::Toeplitz(c(0.7^rep(1:p-1)))
OmegaTrue     <- pracma::inv(SigTrue)
# Generate expected value vector:
mu            <- rep(0,p)
# Generate multivariate normal distribution:
set.seed(123)
X             <- MASS::mvrnorm(n, mu = mu, Sigma = SigTrue)
posterior     <- blockBGL(X, iterations = 1000, burnin = 500, lambda = 0.5)
```

`blockBGR` *Bayesian graphical ridge block Gibbs sampler for Gaussian graphical models.*

Description

Implements a Bayesian graphical ridge block Gibbs sampler to simulate the posterior distribution of the precision matrix for Gaussian graphical models..

© University of Pretoria

Usage

```
blockBGR(X, burnin, iterations, sig = 1, verbose = TRUE)
```

Arguments

`X`	A numeric matrix.
`burnin`	An integer representing the number of burn-in iterations.
`iterations`	An integer representing the length of the Markov chain post burn-in.
`sig`	A numeric value representing the standard deviation parameter for the double Gaussian and truncated Gaussian prior.
`verbose`	A logical indicating if the MCMC sampler progress should be printed.

Value

A list containing precision ‘Omega‘ and covariance ‘Sigma‘ matrices from the Markov chains.

Examples

```
# Generate true covariance matrix:
p             <- 10
n             <- 50
SigTrue       <- pracma::Toeplitz(c(0.7^rep(1:p-1)))
OmegaTrue     <- pracma::inv(SigTrue)
# Generate expected value vector:
mu            <- rep(0,p)
# Generate multivariate normal distribution:
set.seed(123)
X             <- MASS::mvrnorm(n, mu = mu, Sigma = SigTrue)
posterior     <- blockBGR(X, iterations = 1000, burnin = 500, sig = 0.5)
```

© University of Pretoria

© University of Pretoria

Appendix F

Derived Publications

The following is a list of the publications derived from this dissertation.

- Smith, J., Arashi, M., and Bekker, A. (2023). The Bayesian graphical ridge estimator for inverse covariance matrices. (Chapter 3 - submitted)
- Smith, J., Bekker, A., and Arashi, M. (2023) A naïve Bayesian graphical elastic net: driving advances in differential network analysis. (Chapter 4 - submitted)
- Smith, J., Arashi, M., and Bekker, A. (2022). Empowering differential networks using Bayesian analysis. PLOS ONE 17(1): e0261193. https://doi.org/10.1371/journal.pone.0261193 (Chapter 5 - published in peer reviewed journal)

© University of Pretoria

© University of Pretoria